



UNIVERSIDAD NACIONAL DE COLOMBIA

# Evaluación de la robustez a observaciones atípicas de modelos lineales generalizados escalar-imagen vía variación total

Juan Camilo Bernal Castillo

Facultad de Ciencias, Departamento de Estadística  
Bogotá, Colombia  
2021

# Evaluación de la robustez a observaciones atípicas de modelos lineales generalizados escalar-imagen vía variación total

Juan Camilo Bernal Castillo

Trabajo final presentado como requisito parcial para optar al título de:  
**Magister en Ciencias-Estadística**

Director:  
Ph.D. en Estadística, Rubén Darío Guevara González

Plan de estudios: Profundización

Universidad Nacional de Colombia  
Facultad de Ciencias, Departamento de Estadística  
Bogotá, Colombia  
2022



Dedicado a

mis padres y mi hermana.  
A mis abuelos Blanca y Nacho (QEPD).

El camino es la recompensa.



# Agradecimientos

Este trabajo va dedicado a mi familia, por todo el amor y apoyo durante tantos años. Agradezco muy especialmente a Angie Solano por creer en mi y acompañarme en este camino. A David Solano por sus enseñanzas y amistad. A Sandra Pastran, pues sin su ayuda no habría sido posible. Muchas gracias a los profesores Rubén Darío Guevara y Sergio Calderón por toda su guía y compromiso. Gracias a mis compañeros y amigos de la carrera de Estadística de la Universidad Nacional de Colombia.





# Resumen

**Título en español:** Evaluación de la robustez a observaciones atípicas de modelos lineales generalizados escalar-imagen vía variación total

La aplicación de cualquier modelo estadístico a partir de un conjunto de datos de estudio no solo implica el planteamiento y posterior estimación e interpretación de los parámetros del mismo, también requiere de un proceso minucioso de validación de supuestos y de análisis de influencia. Este último es importante para la identificación de unidades muestrales que pueden afectar de forma importante la estimación de los parámetros o que pueden manifestar alguna característica particular dentro de la población objeto de estudio. La identificación de observaciones de alto *leverage*, *outliers* y observaciones influyentes se ha estudiado en profundidad dentro de los modelos lineales clásicos. Sin embargo, la literatura en este tema es escasa dentro de los modelos con datos funcionales. Así, el objetivo de este trabajo es evaluar el nivel de robustez a observaciones atípicas de los modelos lineales generalizados con respuesta escalar y cuyas covariables pueden ser escalares e imágenes, en los cuales los parámetros son estimados vía variación total, por medio de un estudio de simulación. Adicionalmente, se proponen métodos para la identificación de posibles observaciones influyentes.

**Palabras clave:** Datos funcionales, Imágenes, Variación total, Modelos lineales generalizados, observaciones Influyentes.

# Abstract

**Título en inglés:** Evaluation of robustness to outliers of scalar-image generalized linear models via total variation

The application of any statistical model from a dataset of interest implies not only the proposal and posterior estimation and parameter's interpretation. Also requires a meticulous process of assumptions validation and influence analysis. The latter is important for the identification of sample units which could affect importantly the estimates or elements which could exhibit a particular feature inside the objective population. The identification of high leverage, outliers and influential observations has been studied deeply in classic linear models. However, the literature on this topic is scant in models with functional data. So, the objective of this work is to evaluate the level of robustness to atypical observations of generalized linear models with scalar response and whose covariates could be scalar and images, which are estimated via total variation, by a simulation study. In addition, we propose methods to identify possible influential observations.

**Keywords:** Functional data, Images, Total variation, Generalized linear models, Influential observations.



# Lista de Figuras

1-1. Curvas espectrométricas de muestras de carne. Conjunto de datos tecator (paquete de R <i>fda</i> ). . . . .	2
1-2. Primera y segunda derivada de curvas espectrométricas de muestras de carne.	2
1-3. Imagen MRI representada como una matriz de pixeles. Tomado de Poldrack et al. (2011). . . . .	5
2-1. Puntos outlier, de alto leverage e influyentes. Tomado de McCullagh and Nelder (1989). . . . .	9
2-2. Firework plot. Tomado de Jang and Anderson-Cook (2015) . . . . .	10
3-1. Ejemplo de estimación del parámetro superficie $\beta$ . Tomado de Ivanescu (2018).	13
3-2. Imagen observada (izquierda) y componentes del gradiente discreto: sentido vertical (centro) y sentido horizontal (derecha). Tomado de Wang et al. (2017).	16
3-3. Replicaciones bootstrap de un conjunto de curvas (Izquierda). Replicaciones por el método bootstrap suavizado de un conjunto de curvas (Derecha). Tomado de Febrero-Bande et al. (2010) . . . . .	21
3-4. FOM de muestra de imágenes MRI y mapas de calor de valores $DO$ de las imágenes 387 y 92. Tomado de Rousseeuw and Hubert (2018). . . . .	25
3-5. Parámetro imagen $\beta$ . . . . .	26
3-6. MS-PLOT. . . . .	29
4-1. Parámetro imagen $\beta$ . . . . .	32
4-2. Primera fila. Parámetro imagen $\beta_0$ . Segunda fila: Parámetro imagen $\beta_1$ (Modelo de contaminación). . . . .	33
4-3. Valores promedio $d_2(\beta_v, \hat{\beta}_v)$ y valores $d_2(\theta, \hat{\theta})$ . Modelo de respuesta continua, bajo niveles de contaminación $\gamma_c$ (%) y factor de contaminación de magnitud $\gamma_b$ . $\theta_0 = 2$ , $n = 100$ . . . . .	34
4-4. Valores promedio $d_2(\beta_v, \hat{\beta}_v)$ y valores $d_2(\theta, \hat{\theta})$ . Modelo de respuesta continua, bajo niveles de contaminación $\gamma_c$ (%) y contaminación de cambio de forma - Tipo 1. $\theta_0 = 2$ , $\beta_0 = \text{Cuadrado}$ . Línea punteada indica valores promedio de $d_2(\beta_v, \hat{\beta}_v)$ y valores de $d_2(\theta, \hat{\theta})$ , bajo el escenario sin contaminación. . . . .	35
4-5. Valores promedio $d_2(\beta_v, \hat{\beta}_v)$ y valores $d_2(\theta, \hat{\theta})$ . Modelo de respuesta continua, bajo niveles de contaminación $\gamma_c$ (%) y contaminación de cambio de forma - Tipo 2. . . . .	35

4-6. Valores promedio $d_2(\beta_v, \hat{\beta}_v)$ y valores $d_2(\theta, \hat{\theta})$ . Modelo de respuesta binaria, bajo niveles de contaminación $\gamma_c$ (%) y factor de contaminación de magnitud $\gamma_b$ . $\theta_0 = 2$ , $n = 100$ . . . . .	36
4-7. Valores promedio $d_2(\beta_v, \hat{\beta}_v)$ y valores $d_2(\theta, \hat{\theta})$ . Modelo de respuesta binaria, bajo niveles de contaminación $\gamma_c$ (%) y contaminación de cambio de forma - Tipo 1. $\theta_0 = 2$ , $\beta_0 = \text{Cuadrado}$ . . . . .	37
4-8. Curva ROC asociada a modelo de clasificación binaria. Tomado de <a href="#">Yang and Berdine (2017)</a> . . . . .	40
4-9. Valores <i>RMSPE</i> promedio. Modelo de respuesta continua, bajo niveles de contaminación $\gamma_c$ (%) y factor de contaminación de magnitud $\gamma_b$ . $\theta_0 = 2$ . Primera fila ( $n = 100$ ), segunda fila ( $n = 170$ ), tercera fila ( $n = 220$ ). . . . .	41
4-10. Valores <i>RMSPE</i> promedio. Modelo de respuesta continua, bajo niveles de contaminación $\gamma_c$ (%) y factor de contaminación de magnitud $\gamma_b$ . $\theta_0 = 2$ . Primera fila ( $\gamma_b = 2$ ), segunda fila ( $\gamma_b = 5$ ). . . . .	42
4-11. Valores <i>ABC</i> promedio. Modelo de respuesta binaria, bajo niveles de contaminación $\gamma_c$ (%) y factor de contaminación de magnitud $\gamma_b$ . $\theta_0 = 2$ . Primera fila ( $n = 100$ ), segunda fila ( $n = 170$ ), tercera fila ( $n = 220$ ). . . . .	42
4-12. Valores <i>ABC</i> promedio. Modelo de respuesta binaria, bajo niveles de contaminación $\gamma_c$ (%) y factor de contaminación de magnitud $\gamma_b$ . $\theta_0 = 2$ . Primera fila ( $\gamma_b = 2$ ), segunda fila ( $\gamma_b = 5$ ). . . . .	44
4-13. Valores <i>RMSPE</i> promedio. Modelo de respuesta continua, bajo niveles de contaminación $\gamma_c$ (%) y contaminación de forma Tipo 1. $\theta_0 = 2$ . Primera columna ( $\beta_0 = \text{T}$ ), segunda columna ( $\beta_0 = \text{Triángulo}$ ), tercera columna ( $\beta_0 = \text{Cuadrado}$ ), cuarta columna ( $\beta_0 = \text{Ajedrez}$ ). Línea punteada indica valor <i>RMSPE</i> promedio bajo el escenario sin contaminación. . . . .	44
4-14. Valores <i>ABC</i> promedio. Modelo de respuesta binaria, bajo niveles de contaminación $\gamma_c$ (%) y contaminación de forma Tipo 1. $\theta_0 = 2$ . Primera columna ( $\beta_0 = \text{T}$ ), segunda columna ( $\beta_0 = \text{Triángulo}$ ), tercera columna ( $\beta_0 = \text{Cuadrado}$ ), cuarta columna ( $\beta_0 = \text{Ajedrez}$ ). Línea punteada indica valor <i>ABC</i> promedio bajo el escenario sin contaminación. . . . .	45
4-15. Valores <i>RMSPE</i> promedio. Modelo de respuesta continua, bajo niveles de contaminación $\gamma_c$ (%) y contaminación de forma Tipo 2 (zoom del 50% en parámetro imagen). . . . .	45
4-16. Tasas de correcta detección ( $v_c$ ) promedio en 100 replicaciones. Modelo de respuesta continua sin intercepto, bajo niveles de contaminación $\gamma_c$ (%) y factor de contaminación de magnitud $\gamma_b$ . Primera fila ( $\gamma_b = 2$ ), segunda fila ( $\gamma_b = 5$ ). $n = 70$ . Gráficas (a) y (e): $\beta_0 = \text{T}$ , (b) y (f): $\beta_0 = \text{Triángulo}$ , (c) y (g): $\beta_0 = \text{Cuadrado}$ , (d) y (h): $\beta_0 = \text{Ajedrez}$ . . . . .	48

4-17. Tasas de falsa detección ( $f_c$ ) promedio en 100 repeticiones. Modelo de respuesta continua sin intercepto, bajo niveles de contaminación $\gamma_c$ (%) y factor de contaminación de magnitud $\gamma_b$ . Primera fila ( $\gamma_b = 2$ ), segunda fila ( $\gamma_b = 5$ ). Gráficas (a) y (e): $\beta_0 = T$ , (b) y (f): $\beta_0 = \text{Triángulo}$ , (c) y (g): $\beta_0 = \text{Cuadrado}$ , (d) y (h): $\beta_0 = \text{Ajedrez}$ . . . . .	48
4-18. Valores $RMSPE$ promedio en 100 repeticiones y valores $RMSPE$ obtenidos al omitir las observaciones identificadas como influyentes por cada método. Modelo de respuesta continua sin intercepto, bajo niveles de contaminación $\gamma_c$ (%) y factor de contaminación de magnitud $\gamma_b$ . Primera fila ( $\gamma_b = 2$ ), segunda fila ( $\gamma_b = 5$ ). Gráficas (a) y (e): $\beta_0 = T$ , (b) y (f): $\beta_0 = \text{Triángulo}$ , (c) y (g): $\beta_0 = \text{Cuadrado}$ , (d) y (h): $\beta_0 = \text{Ajedrez}$ . . . . .	49
4-19. Primera fila: Tasas de correcta detección ( $v_c$ ) promedio. Segunda fila: Tasas de falsa detección ( $f_c$ ) promedio. Modelo de respuesta continua sin intercepto con $\beta_0 = T$ , bajo niveles de contaminación $\gamma_c$ (%) y contaminación de forma Tipo 1. Gráficas (a) y (d): $\beta_1 = \text{Triángulo}$ , (b) y (e): $\beta_1 = \text{Cuadrado}$ , (c) y (f): $\beta_1 = \text{Ajedrez}$ . . . . .	50
4-20. Tasas de correcta detección ( $v_c$ ) promedio en 100 repeticiones. Modelo de respuesta binaria sin intercepto, bajo niveles de contaminación $\gamma_c$ (%) y factor de contaminación de magnitud $\gamma_b$ . $n = 70$ . Primera fila ( $\gamma_b = 2$ ), segunda fila ( $\gamma_b = 5$ ). Gráficas (a) y (e): $\beta_0 = T$ , (b) y (f): $\beta_0 = \text{Triángulo}$ , (c) y (g): $\beta_0 = \text{Cuadrado}$ , (d) y (h): $\beta_0 = \text{Ajedrez}$ . . . . .	51
4-21. Tasas de falsa detección ( $f_c$ ) promedio en 100 repeticiones. Modelo de respuesta binaria sin intercepto, bajo niveles de contaminación $\gamma_c$ (%) y factor de contaminación de magnitud $\gamma_b$ . $n = 70$ . Primera fila ( $\gamma_b = 2$ ), segunda fila ( $\gamma_b = 5$ ). Gráficas (a) y (e): $\beta_0 = T$ , (b) y (f): $\beta_0 = \text{Triángulo}$ , (c) y (g): $\beta_0 = \text{Cuadrado}$ , (d) y (h): $\beta_0 = \text{Ajedrez}$ . . . . .	51
4-22. Primera fila: Tasas de correcta detección ( $v_c$ ) promedio. Segunda fila: Tasas de falsa detección ( $f_c$ ) promedio. Modelo de respuesta binaria sin intercepto con $\beta_0 = T$ , bajo niveles de contaminación $\gamma_c$ (%) y contaminación de forma Tipo 1. Gráficas (a) y (d): $\beta_1 = \text{Triángulo}$ , (b) y (e): $\beta_1 = \text{Cuadrado}$ , (c) y (f): $\beta_1 = \text{Ajedrez}$ . . . . .	52
4-23. Primera fila: Tasas de correcta detección ( $v_c$ ) promedio. Segunda fila: Tasas de falsa detección ( $f_c$ ) promedio. Modelo de respuesta continua sin intercepto con $\beta_0 = T$ , bajo niveles de contaminación $\gamma_c$ (%) y contaminación de forma Tipo 2 (zoom del 50% en parámetro imagen). Gráficas (a) y (e): $\beta_0 = T$ , (b) y (f): $\beta_0 = \text{Triángulo}$ , (c) y (g): $\beta_0 = \text{Cuadrado}$ , (d) y (h): $\beta_0 = \text{Ajedrez}$ . . .	52
4-24. Ejemplo MRI de paciente sano (izquierda) y paciente con tumor cerebral (derecha). . . . .	53
4-25. Estimación del parámetro $\beta$ . . . . .	54

---

4-26. Residuos de Pearson (Izquierda), Residuos componente de desvío (Centro) y Bandas del 95 % de confianza de los residuos componente de desvío (Derecha).	55
4-27. Imágenes MRI de unidades muestrales 94, 107 y 129 de la muestra de entrenamiento. . . . .	55
4-28. Curva ROC asociada al modelo estimado. . . . .	56
4-29. FOM y MS-PLOT de estimaciones de $\hat{\beta}_{(i)}$ , $i = 1, \dots, 202$ . . . . .	56
4-30. Imágenes MRI de unidades muestrales 53 (a) y 136 (c) y mapas de calor de valores DO de estimaciones $\hat{\beta}_{(53)}$ (b) y $\hat{\beta}_{(136)}$ (d). . . . .	58

# Lista de Tablas

<b>3-1.</b> Proporción de veces que se rechaza la hipótesis nula de normalidad en 1000 repeticiones, bajo niveles de significancia $\alpha = 0.05$ y $\alpha = 0.1$ , usando las pruebas de normalidad Shapiro-Wilk, Jarque-Bera y Kolmogorov-Smirnov. . . . .	27
<b>4-1.</b> Matriz de confusión para un modelo de clasificación binaria . . . . .	39
<b>4-2.</b> Matriz de confusión de datos de prueba. . . . .	56
<b>4-3.</b> Matriz de confusión de datos de prueba, omitiendo la unidad muestral 136 de conjunto de entrenamiento. . . . .	58
<b>A-1.</b> Promedio y desviación estándar (en paréntesis) de $RMPSE$ , $d(\beta, \hat{\beta})$ , $d_2(\widehat{\beta}_v, \widehat{\beta}_v)$ , $d(\theta, \hat{\theta})$ , $d_2(\theta, \hat{\theta})$ , $\hat{\theta}$ y tiempo total de simulación de 200 repeticiones. Modelo de respuesta continua, $n = 100$ , bajo niveles de contaminación $\gamma_c$ y factor de contaminación de magnitud $\gamma_b$ . . . . .	63
<b>A-2.</b> Promedio y desviación estándar (en paréntesis) de $RMPSE$ , $d(\beta, \hat{\beta})$ , $d_2(\widehat{\beta}_v, \widehat{\beta}_v)$ , $d(\theta, \hat{\theta})$ , $d_2(\theta, \hat{\theta})$ , $\hat{\theta}$ y tiempo total de simulación de 200 repeticiones. Modelo de respuesta continua, $n = 170$ , bajo niveles de contaminación $\gamma_c$ y factor de contaminación de magnitud $\gamma_b$ . . . . .	64
<b>A-3.</b> Promedio y desviación estándar (en paréntesis) de $RMPSE$ , $d(\beta, \hat{\beta})$ , $d_2(\widehat{\beta}_v, \widehat{\beta}_v)$ , $d(\theta, \hat{\theta})$ , $d_2(\theta, \hat{\theta})$ , $\hat{\theta}$ y tiempo total de simulación de 200 repeticiones. Modelo de respuesta continua, $n = 220$ , bajo niveles de contaminación $\gamma_c$ y factor de contaminación de magnitud $\gamma_b$ . . . . .	65
<b>A-4.</b> Promedio y desviación estándar (en paréntesis) de $ABC$ , $d(\beta, \hat{\beta})$ , $d_2(\widehat{\beta}_v, \widehat{\beta}_v)$ , $d(\theta, \hat{\theta})$ , $d_2(\theta, \hat{\theta})$ , $\hat{\theta}$ y tiempo total de simulación de 200 repeticiones. Modelo de respuesta binaria, $n = 100$ , bajo niveles de contaminación $\gamma_c$ y factor de contaminación de magnitud $\gamma_b$ . . . . .	66
<b>A-5.</b> Promedio y desviación estándar (en paréntesis) de $ABC$ , $d(\beta, \hat{\beta})$ , $d_2(\widehat{\beta}_v, \widehat{\beta}_v)$ , $d(\theta, \hat{\theta})$ , $d_2(\theta, \hat{\theta})$ , $\hat{\theta}$ y tiempo total de simulación de 200 repeticiones. Modelo de respuesta binaria, $n = 170$ , bajo niveles de contaminación $\gamma_c$ y factor de contaminación de magnitud $\gamma_b$ . . . . .	67
<b>A-6.</b> Promedio y desviación estándar (en paréntesis) de $ABC$ , $d(\beta, \hat{\beta})$ , $d_2(\widehat{\beta}_v, \widehat{\beta}_v)$ , $d(\theta, \hat{\theta})$ , $d_2(\theta, \hat{\theta})$ , $\hat{\theta}$ y tiempo total de simulación de 200 repeticiones. Modelo de respuesta binaria, $n = 220$ , bajo niveles de contaminación $\gamma_c$ y factor de contaminación de magnitud $\gamma_b$ . . . . .	68

<b>A-7.</b> Promedio y desviación estándar (en paréntesis) de $RMSPE$ , $d(\beta, \hat{\beta})$ , $d_2(\widehat{\beta}_v, \beta_v)$ , $d(\theta, \hat{\theta})$ , $d_2(\theta, \hat{\theta})$ , $\hat{\theta}$ y tiempo total de simulación de 200 repeticiones. Modelo de respuesta continua, $n = 100$ , bajo niveles de contaminación $\gamma_c$ y contaminación de cambio de forma - Tipo 1 en parámetro imagen. . . . .	69
<b>A-8.</b> Promedio y desviación estándar (en paréntesis) de $RMSPE$ , $d(\beta, \hat{\beta})$ , $d_2(\widehat{\beta}_v, \beta_v)$ , $d(\theta, \hat{\theta})$ , $d_2(\theta, \hat{\theta})$ , $\hat{\theta}$ y tiempo total de simulación de 200 repeticiones. Modelo de respuesta continua, $n = 170$ , bajo niveles de contaminación $\gamma_c$ y contaminación de cambio de forma - Tipo 1 en parámetro imagen. . . . .	70
<b>A-9.</b> Promedio y desviación estándar (en paréntesis) de $RMSPE$ , $d(\beta, \hat{\beta})$ , $d_2(\widehat{\beta}_v, \beta_v)$ , $d(\theta, \hat{\theta})$ , $d_2(\theta, \hat{\theta})$ , $\hat{\theta}$ y tiempo total de simulación de 200 repeticiones. Modelo de respuesta continua, $n = 220$ , bajo niveles de contaminación $\gamma_c$ y contaminación de cambio de forma - Tipo 1 en parámetro imagen. . . . .	71
<b>A-10</b> Promedio y desviación estándar (en paréntesis) de $ABC$ , $d(\beta, \hat{\beta})$ , $d_2(\widehat{\beta}_v, \beta_v)$ , $d(\theta, \hat{\theta})$ , $d_2(\theta, \hat{\theta})$ , $\hat{\theta}$ y tiempo total de simulación de 200 repeticiones. Modelo de respuesta binaria, $n = 100$ , bajo niveles de contaminación $\gamma_c$ y contaminación de cambio de forma - Tipo 1 en parámetro imagen. . . . .	72
<b>A-11</b> Promedio y desviación estándar (en paréntesis) de $ABC$ , $d(\beta, \hat{\beta})$ , $d_2(\widehat{\beta}_v, \beta_v)$ , $d(\theta, \hat{\theta})$ , $d_2(\theta, \hat{\theta})$ , $\hat{\theta}$ y tiempo total de simulación de 200 repeticiones. Modelo de respuesta binaria, $n = 170$ , bajo niveles de contaminación $\gamma_c$ y contaminación de cambio de forma - Tipo 1 en parámetro imagen. . . . .	73
<b>A-12</b> Promedio y desviación estándar (en paréntesis) de $ABC$ , $d(\beta, \hat{\beta})$ , $d_2(\widehat{\beta}_v, \beta_v)$ , $d(\theta, \hat{\theta})$ , $d_2(\theta, \hat{\theta})$ , $\hat{\theta}$ y tiempo total de simulación de 200 repeticiones. Modelo de respuesta binaria, $n = 220$ , bajo niveles de contaminación $\gamma_c$ y contaminación de cambio de forma - Tipo 1 en parámetro imagen. . . . .	74
<b>A-13</b> Promedio y desviación estándar (en paréntesis) de $RMSPE$ , $d(\beta, \hat{\beta})$ , $d_2(\widehat{\beta}_v, \beta_v)$ , $d(\theta, \hat{\theta})$ , $d_2(\theta, \hat{\theta})$ , $\hat{\theta}$ y tiempo total de simulación de 200 repeticiones. Modelo de respuesta continua, $n = 100$ , bajo niveles de contaminación $\gamma_c$ y contaminación de cambio de forma - Tipo 2 en parámetro imagen (zoom 50 % de parámetro imagen original). . . . .	75
<b>A-14</b> Promedio y desviación estándar (en paréntesis) de $RMSPE$ , $d(\beta, \hat{\beta})$ , $d_2(\widehat{\beta}_v, \beta_v)$ , $d(\theta, \hat{\theta})$ , $d_2(\theta, \hat{\theta})$ , $\hat{\theta}$ y tiempo total de simulación de 200 repeticiones. Modelo de respuesta continua, $n = 170$ , bajo niveles de contaminación $\gamma_c$ y contaminación de cambio de forma - Tipo 2 en parámetro imagen (zoom 50 % de parámetro imagen original). . . . .	75
<b>A-15</b> Promedio y desviación estándar (en paréntesis) de $RMSPE$ , $d(\beta, \hat{\beta})$ , $d_2(\widehat{\beta}_v, \beta_v)$ , $d(\theta, \hat{\theta})$ , $d_2(\theta, \hat{\theta})$ , $\hat{\theta}$ y tiempo total de simulación de 200 repeticiones. Modelo de respuesta continua, $n = 220$ , bajo niveles de contaminación $\gamma_c$ y contaminación de cambio de forma - Tipo 2 en parámetro imagen (zoom 50 % de parámetro imagen original). . . . .	75



<b>A-16</b> Promedio y desviación estándar (en paréntesis) de $RMSPE$ , tasa de correcta detección ( $v_c$ ), tasa de falsa detección ( $f_c$ ) de cada método de identificación de influentes, en 100 replicaciones. Modelo de respuesta continua, $n = 50$ , bajo niveles de contaminación $\gamma_c$ y factor de contaminación de magnitud $\gamma_b$ .	76
<b>A-17</b> Promedio y desviación estándar (en paréntesis) de $RMSPE$ , tasa de correcta detección ( $v_c$ ), tasa de falsa detección ( $f_c$ ) de cada método de identificación de influentes, en 100 replicaciones. Modelo de respuesta continua, $n = 70$ , bajo niveles de contaminación $\gamma_c$ y factor de contaminación de magnitud $\gamma_b$ .	77
<b>A-18</b> Promedio y desviación estándar (en paréntesis) de $ABC$ , tasa de correcta detección ( $v_c$ ), tasa de falsa detección ( $f_c$ ) de cada método de identificación de influentes, en 100 replicaciones. Modelo de respuesta binaria, $n = 50$ , bajo niveles de contaminación $\gamma_c$ y factor de contaminación de magnitud $\gamma_b$ . . . .	78
<b>A-19</b> Promedio y desviación estándar (en paréntesis) de $ABC$ , tasa de correcta detección ( $v_c$ ), tasa de falsa detección ( $f_c$ ) de cada método de identificación de influentes, en 100 replicaciones. Modelo de respuesta binaria, $n = 70$ , bajo niveles de contaminación $\gamma_c$ y factor de contaminación de magnitud $\gamma_b$ . . . .	79
<b>A-20</b> Promedio y desviación estándar (en paréntesis) de $RMSPE$ , tasa de correcta detección ( $v_c$ ), tasa de falsa detección ( $f_c$ ) de cada método de identificación de influentes, en 100 replicaciones. Modelo de respuesta continua, $n = 50$ , bajo de contaminación de cambio forma - Tipo 1 en parámetro imagen. . . .	80
<b>A-21</b> Promedio y desviación estándar (en paréntesis) de $RMSPE$ , tasa de correcta detección ( $v_c$ ), tasa de falsa detección ( $f_c$ ) de cada método de identificación de influentes, en 100 replicaciones. Modelo de respuesta continua, $n = 70$ , bajo de contaminación de cambio forma - Tipo 1 en parámetro imagen. . . .	81
<b>A-22</b> Promedio y desviación estándar (en paréntesis) de $ABC$ , tasa de correcta detección ( $v_c$ ), tasa de falsa detección ( $f_c$ ) de cada método de identificación de influentes, en 100 replicaciones. Modelo de respuesta binaria, $n = 50$ , bajo de contaminación de cambio forma - Tipo 1 en parámetro imagen. . . . .	82
<b>A-23</b> Promedio y desviación estándar (en paréntesis) de $ABC$ , tasa de correcta detección ( $v_c$ ), tasa de falsa detección ( $f_c$ ) de cada método de identificación de influentes, en 100 replicaciones. Modelo de respuesta binaria, $n = 70$ , bajo de contaminación de cambio forma - Tipo 1 en parámetro imagen. . . . .	83
<b>A-24</b> Promedio y desviación estándar (en paréntesis) de $RMSPE$ , tasa de correcta detección ( $v_c$ ), tasa de falsa detección ( $f_c$ ) de cada método de identificación de influentes, en 100 replicaciones. Modelo de respuesta continua, $n = 50$ , bajo de contaminación de cambio forma - Tipo 2 en parámetro imagen (zoom 50 % de parámetro imagen original). . . . .	84

---

<b>A-25</b> Promedio y desviación estándar (en paréntesis) de $RMSPE$ , tasa de correcta detección ( $v_c$ ), tasa de falsa detección ( $f_c$ ) de cada método de identificación de influyentes, en 100 replicaciones. Modelo de respuesta continua, $n = 70$ , bajo de contaminación de cambio forma - Tipo 2 en parámetro imagen (zoom 50% de parámetro imagen original). . . . .	85
---	----

# Contenido

<b>Agradecimientos</b>	<b>VII</b>
<b>Resumen</b>	<b>IX</b>
<b>Lista de figuras</b>	<b>XI</b>
<b>Lista de tablas</b>	<b>XV</b>
<b>1. Introducción</b>	<b>1</b>
<b>2. Modelos Lineales Generalizados</b>	<b>6</b>
2.1. Familia Exponencial Lineal . . . . .	6
2.2. Estimación de Parámetros . . . . .	7
2.3. Análisis de Influencia . . . . .	9
<b>3. Análisis de Datos Funcionales</b>	<b>11</b>
3.1. Modelos con datos funcionales . . . . .	12
3.2. MLG escalar-función estimado via variación total . . . . .	14
3.3. Medidas de influencia en modelos escalar-función . . . . .	19
3.4. Detección de outliers en conjuntos de imágenes . . . . .	21
3.4.1. Directional Outlyingness . . . . .	23
3.4.2. MS-PLOT . . . . .	27
<b>4. Evaluación de la robustez a observaciones atípicas de modelos GSIRM-TV</b>	<b>31</b>
4.1. Calidad de estimaciones de modelos GSIRM-TV ante presencia de observa- ciones atípicas . . . . .	31
4.2. Predicción de nuevas observaciones en modelos GSIRM-TV ante presencia de observaciones atípicas . . . . .	37
4.3. Identificación de observaciones influyentes en modelos GSIRM-TV . . . . .	46
4.4. Aplicación en datos reales . . . . .	53
<b>5. Conclusiones y recomendaciones</b>	<b>59</b>
5.1. Conclusiones . . . . .	59
5.2. Recomendaciones . . . . .	60

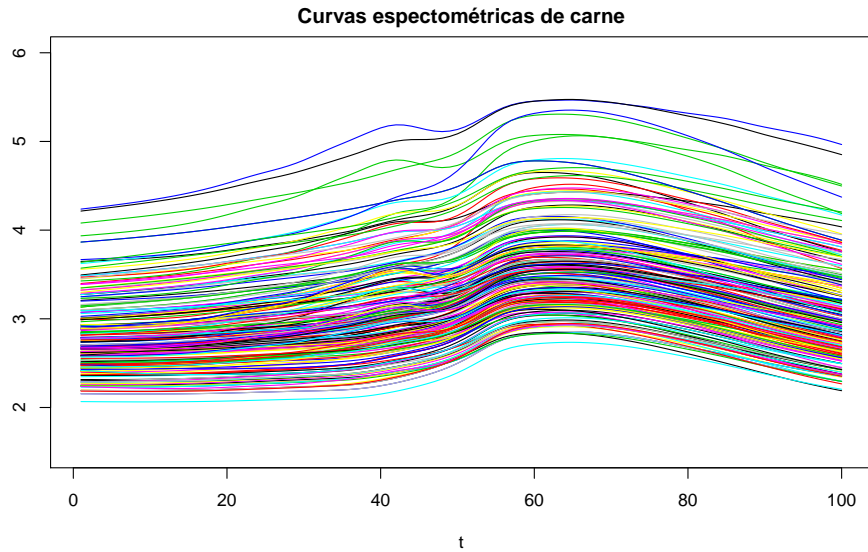
---

<b>A. Anexo: Resultados Estudio de Simulación</b>	<b>62</b>
<b>Bibliografía</b>	<b>86</b>

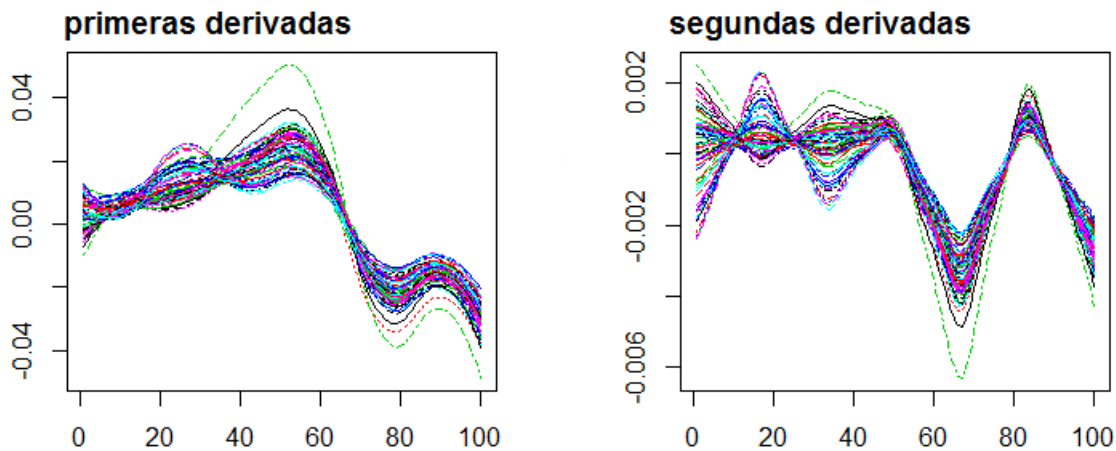
# 1. Introducción

El Análisis de Datos Funcionales o FDA por sus siglas en inglés (*Functional Data Analysis*) es un área de la Estadística relativamente nueva, que se ha desarrollado ampliamente en las últimas dos décadas. En este campo es de interés el estudio de objetos que pueden ser vistos como funciones con dominio univariado o multivariado y cuyo conjunto de llegada puede ser la recta real o  $\mathbb{R}^d$  ( $d \geq 1$ ). De forma general, en FDA es posible estudiar unidades observacionales que provienen de un proceso aleatorio  $\mathbf{X}(\mathbf{t}) = (X^{(1)}(t_1), \dots, X^{(p)}(t_p))$ ,  $\mathbf{t} := (t_1, t_2, \dots, t_p) \in \Omega := \Omega_1 \times \dots \times \Omega_p$ , con funciones aleatorias  $X^{(j)}(t_j) : \Omega_j \rightarrow \mathbb{R}$ ,  $\Omega_j \subset \mathbb{R}^{d_j}$ ,  $d_j \in \mathbb{N}$ ,  $j = 1, \dots, p$ , donde los dominios de las funciones,  $\Omega_j$ , no son iguales necesariamente. Es decir, el vector aleatorio  $\mathbf{X}(\mathbf{t})$ , puede estar compuesto por funciones de una variable o de varias variables, simultáneamente. A modo de ejemplo, observe la Figura (1-1). Esta Figura corresponde a la gráfica de curvas espectrométricas medidas en muestras de carne. El estudio de este tipo de objetos es frecuente en áreas relacionadas con la química, donde el objetivo es determinar las propiedades del elemento de interés. En este caso,  $p = 1$  y  $d = 1$ , de modo que el proceso aleatorio asociado a los datos de interés se puede definir como  $X(t) : \Omega \rightarrow \mathbb{R}$ ,  $\Omega \subset \mathbb{R}$ .

Esta generalización permite el estudio de imágenes en 2 dimensiones, en cuyo caso el proceso asociado puede definirse como  $X(\mathbf{t}) : \Omega \rightarrow \mathbb{R}$ ,  $\Omega \subset \mathbb{R}^2$  ( $p = 1$ ,  $d = 2$ ), o podemos considerar procesos funcionales multivariados. Para el ejemplo de la Figura (1-1), si tuviésemos en cuenta las primeras dos derivadas de las funciones, las cuales pueden aportar información importante del proceso, se obtendría el proceso funcional multivariado  $\mathbf{X}(t) = (X^{(1)}(t), X^{(2)}(t), X^{(3)}(t))$ , siendo  $X^{(2)} = [X^{(1)}]'$ ,  $X^{(3)} = [X^{(1)}]''$ . En este caso donde  $p = 3$  y  $d_j = 1$  para todo  $j = 1, 2, 3$  (ver Figura (1-2)).



**Figura 1-1.:** Curvas espectrométricas de muestras de carne. Conjunto de datos tecator (paquete de R *fda*).



**Figura 1-2.:** Primera y segunda derivada de curvas espectrométricas de muestras de carne.

El Análisis de Datos Funcionales tiene múltiples aplicaciones en áreas como la química, ingeniería, finanzas, entre muchas otras áreas. En la medicina, por ejemplo, en el estudio de enfermedades asociadas con el funcionamiento cerebral, es de interés la identificación de marcadores que ayuden a predecir resultados clínicos. Dentro de estos marcadores podemos encontrar test cognitivos, diseñados para identificar ciertas anomalías del funcionamiento del cerebro o marcadores genéticos. Hoy en día, gracias a grandes avances tecnológicos en materia de captura y procesamiento de imágenes de resonancia magnética, es posible analizar información obtenida a partir de estos conjuntos de imágenes. Más aún, desde el FDA se han propuesto algunos modelos lineales para determinar la relación entre las imágenes

capturadas vía tomografías computarizadas, con algún resultado clínico como por ejemplo si el paciente padece Alzheimer o no, o algún otro detrimento de las funciones cerebrales.

Dicho lo anterior, en los modelos estadísticos propuestos en este campo suponemos una variable respuesta escalar  $Y$ , y un conjunto de covariables que bien pueden ser cualitativas o cuantitativas (ej, género y edad del paciente) y una covariable adicional que corresponde a una imagen, que puede ser la resonancia magnética del cerebro de la persona, también conocida como imagen MRI por sus siglas en inglés (*Magnetic Resonance Image*), (ver Figura (1-3)). Wang et al. (2017) proponen el siguiente modelo lineal generalizado, denominado GSIRM-TV, (*Generalized Scalar-on-Image Regression Models via Total Variation*):

Se consideran un conjunto de predictores  $Z \in \mathbb{R}^p$  y  $X$  una imagen que pertenece a un espacio de dimensión infinita. Se asume que:

$$\begin{aligned}
 Y|(X, Z) &\sim \text{Familia Exponencial Lineal } (\mu, \phi) \\
 g(\mu) &= \theta^t Z + \langle X, \beta_0 \rangle,
 \end{aligned}
 \tag{1-1}$$

donde  $\mu$  es el parámetro de localización,  $\phi$  el parámetro de dispersión y  $g(\cdot)$  la función de enlace. A partir de los datos observados, se desea estimar el vector de parámetros  $\theta = (\theta_1, \dots, \theta_p)^t$  y el parámetro imagen  $\beta_0$ . Dentro de los modelos lineales con respuesta escalar y covariable funcional con dominio en  $\mathbb{R}^2$  (imágenes bidimensionales), el modelo descrito en (1-1) se diferencia de otras propuestas al considerar que el parámetro  $\beta_0$  pertenece al espacio de variación acotada  $BV$  (*Bounded Variation*). En general, en la teoría de datos funcionales se supone que un proceso estocástico  $X$  proviene del espacio  $L^2$  de funciones cuadrado integrables [Horváth and Kokoszka (2012)], el cual es un espacio adecuado para funciones suaves. Sin embargo, Wang et al. (2017) argumentan que en las imágenes se observan saltos y bordes de distintas formas que los métodos de suavizado utilizados al trabajar sobre el espacio  $L^2$  no capturan de forma adecuada. Así, el modelo lineal generalizado descrito es estimado a partir de métodos de regularización basados en variación total.

Por otra parte, en general, una vez se estima cualquier tipo de modelo estadístico, clásico o funcional, es necesario llevar a cabo el análisis diagnóstico del mismo donde se verifica el cumplimiento de los supuestos del modelo. Esta parte del modelamiento es fundamental para verificar que la inferencia realizada a partir del modelo estimado (prueba de hipótesis, intervalos de confianza y predicciones) sea correcta. Una componente muy importante dentro de ese análisis diagnóstico es la identificación de observaciones influyentes que puedan afectar seriamente las estimaciones de los parámetros del modelo. Estas observaciones pueden corresponder a individuos que presentan valores atípicos en la variable respuesta (*outliers*), o cuyas variables predictoras están alejadas del espacio de las covariables, también conocidos

como puntos de alto *leverage* [McCullagh and Nelder (1989)].

Considere un modelo lineal generalizado clásico donde la media del proceso estocástico  $Y$  se puede modelar a partir de un conjunto de covariables escalares  $x_1, \dots, x_p$ . Esto es, suponga  $Y_1, Y_2, \dots, Y_n$  un conjunto de variables aleatorias tales que

$$\begin{cases} Y_k \sim \text{Familia Exponencial Lineal } (\mu_k, \phi_k) \\ g(\mu_k) = \beta_1 + \beta_2 x_{2k} + \dots + \beta_p x_{pk} \\ Y_1, \dots, Y_n \text{ independientes,} \end{cases} \quad (1-2)$$

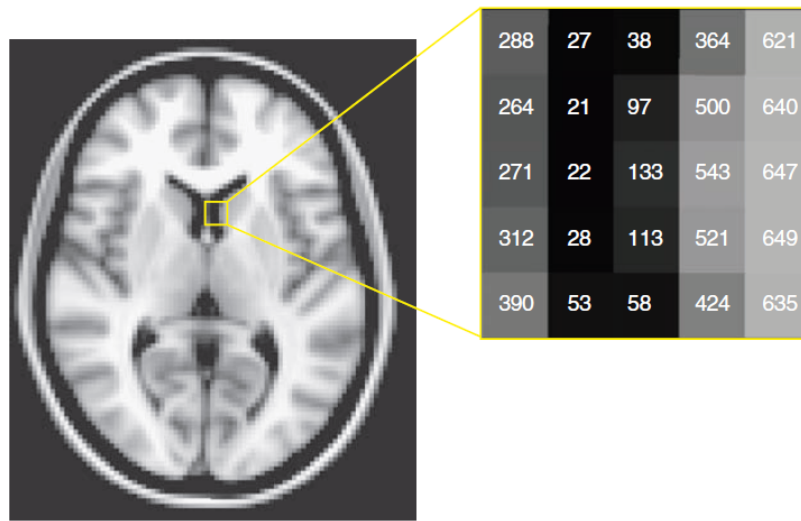
donde  $\mu_k = E(Y_k)$  y  $g(\cdot)$  es la función de enlace. Una vez ajustado el modelo a partir del conjunto de datos observados, en el análisis de influencia del modelo la medida más utilizada para determinar si una observación es influyente es la distancia de Cook [Cook (1977)]. Esta medida se define a continuación para el individuo  $k$ -ésimo:

$$DC_k = \left( \hat{\beta}_{(k)} - \hat{\beta} \right)^t \widehat{Var}(\hat{\beta})^{-1} \left( \hat{\beta}_{(k)} - \hat{\beta} \right) \quad (1-3)$$

donde  $\hat{\beta}$  es el vector de parámetros estimado con todas las observaciones y  $\hat{\beta}_{(k)}$  es la estimación al eliminar al individuo  $k$ . Intuitivamente, se dice que una observación es influyente si la distancia entre el vector de parámetros estimado y la estimación al eliminar dicho individuo es grande. La distancia de Cook se basa en la distancia de Mahalanobis, la cual se diferencia de la distancia Euclidiana entre dos vectores en que tiene en cuenta la correlación entre las estimaciones de las componentes del vector  $\beta$ .

Ahora, en el caso del modelo lineal generalizado escalar-imagen presentado en la expresión (1-1), no existe en la literatura una medida adecuada para identificar observaciones influyentes en la estimación de los parámetros del modelo. Más aún, es necesario investigar qué tanto pueden verse afectadas las estimaciones del parámetro imagen  $\beta_0$ , y de los parámetros escalares  $\theta_j$ ,  $j = 1, \dots, p$ , ante la presencia de observaciones atípicas. Este estudio es relevante, dado que a diferencia de los modelos lineales clásicos, una observación particular puede tener alta influencia en la estimación de parámetros (escalares), mientras que en modelos escalar-imagen intuitivamente es difícil pensar que una sola observación pueda afectar la estimación del parámetro imagen en una gran parte de su dominio. Esto en la práctica se traduce en afectación de un alto número de píxeles. Por tal razón, el objetivo principal de este trabajo es evaluar el nivel de robustez a observaciones atípicas de los modelos lineales generalizados con respuesta escalar y cuyas covariables pueden incluir variables escalares e imágenes, estimados vía variación total. Esta evaluación se hará a través de estudio de simulación para determinar el nivel de afectación tanto en las estimaciones del parámetro imagen  $\beta$  y los parámetros escalares  $\theta$ , como en el desempeño del modelo en términos de predicción, bajo diferentes escenarios de contaminación en la muestra.





**Figura 1-3.:** Imagen MRI representada como una matriz de pixeles. Tomado de [Poldrack et al. \(2011\)](#).

## 2. Modelos Lineales Generalizados

Los Modelos Lineales Generalizados (MLG) son una extensión de los modelos de regresión lineal múltiple para el caso en el que la variable respuesta no sigue una distribución normal dado un conjunto de covariables  $x_1, \dots, x_p$ . Por lo tanto, constituyen una herramienta de gran utilidad para modelar datos donde la variable respuesta es binaria (éxito o fracaso), conteos o una variable continua que toma valores estrictamente mayores a 0 y que no pueden ser modelados de forma adecuada a partir de la teoría normal de los modelos de regresión clásicos.

### 2.1. Familia Exponencial Lineal

Retomando la expresión (1-2), consideremos un conjunto de variables aleatorias  $Y_1, Y_2, \dots, Y_n$  tales que

$$\begin{cases} Y_k \sim \text{Familia Exponencial Lineal } (\mu_k, \phi_k) \\ g(\mu_k) = \beta_1 + \beta_2 x_{2k} + \dots + \beta_p x_{pk} \\ Y_1, \dots, Y_n \text{ independientes} \end{cases}$$

donde  $\mu_k$  es la media de la variable aleatoria  $Y_k$ ,  $\phi_k$  se le conoce como parámetro de dispersión,  $g(\cdot)$  es la función de enlace y  $x_1, \dots, x_p$  son las covariables escalares. Si  $Y$  provienen de una distribución que pertenece a la Familia Exponencial Lineal (FEL), entonces su función de densidad se puede escribir de la siguiente forma:

$$f(y, \mu, \phi) = \exp\left(\frac{y\eta - b(\eta)}{a(\phi)} + c(\phi, y)\right), \quad (2-1)$$

donde  $a(\cdot)$ ,  $b(\cdot)$  y  $c(\cdot)$ , son funciones específicas, y  $\eta(\cdot)$  es una función que  $\mu$ . Algunas distribuciones conocidas que pertenecen a esta familia de distribuciones son la distribución Normal, Binomial, Poisson y Gamma.

**Ejemplo 1.** Supongamos que  $Y \sim \text{Ber}(\mu)$ . Entonces, sea  $f_Y(y)$  su función de densidad,

$$\begin{aligned} f_Y(y) &= \mu^y(1-\mu)^{1-y} = \exp(\ln(\mu^y(1-\mu)^{1-y})) \\ &= \exp(y \ln(\mu) + (1-y) \ln(1-\mu)) = \exp(y[\ln(\mu) - \ln(1-\mu)] + \ln(1-\mu)) \\ &= \exp\left(y \ln\left(\frac{\mu}{1-\mu}\right) - (-1) \ln(1-\mu)\right) = \exp\left(\frac{y\eta - b(\eta)}{a(\psi)} + c(\psi, y)\right) \end{aligned}$$

$$\text{Así, } \eta = \ln\left(\frac{\mu}{1-\mu}\right), b(\eta) = -\ln(1-\mu), a(\psi) = 1 \text{ y } c(\psi, y) = 0$$

Para el caso de la distribución normal de parámetros  $\mu$  y  $\sigma^2$ , por ejemplo, se puede verificar que la función de densidad se puede expresar como en la expresión (2-1) con  $\eta = \mu$ ,  $\psi = \sigma^2$ ,  $b(\eta) = \eta^2/2$ ,  $a(\psi) = \psi$  y  $c(y, \psi) = -1/2(y^2/\psi + \ln(2\pi\sigma^2))$ .

## 2.2. Estimación de Parámetros

Consideremos el siguiente modelo de regresión lineal múltiple  $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$ , donde  $\mathbf{Y} = (y_1, \dots, y_n)^t$ ,  $\mathbf{e} = (e_1, \dots, e_n)^t$ ,  $\boldsymbol{\beta} = (\beta_0, \dots, \beta_p)^t$  y  $\mathbf{X}$  la matriz diseño asociada. Supongamos que  $\mathbf{e} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ . Bajo la teoría de los modelos lineales, se puede mostrar que el estimador de mínimos cuadrados ordinarios y el estimados máximo verosimil de  $\boldsymbol{\beta}$  coinciden y que además este estimador se puede obtener de forma cerrada. Esto es  $\hat{\boldsymbol{\beta}} = (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{Y}$  [Hocking (2013)].

Por otro lado, la estimación usual en MLG se lleva a cabo por el método de máxima verosimilitud. Sin embargo, para modelos distintos al normal, no es posible obtener el estimador máximo verosimil de forma analítica, y se requiere el uso de métodos numéricos para encontrar la solución al siguiente problema de optimización:

Bajo el modelo (1-2), dado el conjunto de datos observados, se busca el valor de  $\boldsymbol{\beta} \in R^p$  que maximice la función de verosimilitud  $L(\boldsymbol{\beta})$  dada por la siguiente expresión:

$$L(\boldsymbol{\beta}) = \prod_{k=1}^n \left\{ \exp \left\{ \phi_k^{-1} [y_k \theta(\mu_k) - b[\theta(\mu_k)]] + c(y, \phi_k) \right\} \right\} . \quad (2-2)$$

Maximizar  $L(\boldsymbol{\beta})$  es equivalente a maximizar la función de log-verosimilitud  $l(\boldsymbol{\beta}) = \ln(L(\boldsymbol{\beta}))$ , que se presenta a continuación:

$$l(\boldsymbol{\beta}) = \sum_{k=1}^n \left\{ \phi_k^{-1} [y_k \theta(\mu_k) - b[\theta(\mu_k)]] + c(y_k, \phi_k) \right\} . \quad (2-3)$$

Derivando en términos de las componentes del vector de parámetros, se obtiene la denominada función de score  $U(\boldsymbol{\beta}) = \left[ \frac{\partial l(\boldsymbol{\beta})}{\partial \beta_1}, \dots, \frac{\partial l(\boldsymbol{\beta})}{\partial \beta_p} \right]^T$ . Se puede demostrar que la  $j$ -ésima componente de la función de score es

$$\frac{\partial l(\boldsymbol{\beta})}{\partial \beta_j} = \frac{1}{\phi} \sum_{k=1}^n \frac{(y_k - \mu_k)}{\sqrt{V(\mu_k)}} \sqrt{w_k} x_{kj}$$

donde  $w_k = \frac{1}{V(\mu_k)} \left[ \frac{\partial \mu_k}{\partial \eta_k} \right]^2$ , siendo  $V(\mu_k) = b''[\theta(\mu_k)]$ .

$$U(\boldsymbol{\beta}) = \frac{1}{\phi} \mathbf{X}^T \mathbf{W}^{1/2} \mathbf{V}^{-1/2} (\mathbf{y} - \boldsymbol{\mu})$$

con  $\mathbf{W} = \text{diag}\{w_1, \dots, w_n\}$  y  $\mathbf{V} = \text{diag}\{V(\mu_1), \dots, V(\mu_n)\}$  [McCullagh and Nelder (1989)]. Usando la expansión en serie de Taylor de la función de score alrededor de un punto  $\boldsymbol{\beta}^{(0)}$ , se obtiene que

$$U(\boldsymbol{\beta}) \approx U(\boldsymbol{\beta}^{(0)}) + \left[ \frac{\partial U(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T} \right]_{\boldsymbol{\beta}=\boldsymbol{\beta}^{(0)}} (\boldsymbol{\beta} - \boldsymbol{\beta}^{(0)})$$

y evaluando en el punto  $\hat{\boldsymbol{\beta}}$ ,

$$U(\hat{\boldsymbol{\beta}}) \approx U(\boldsymbol{\beta}^{(0)}) + \left[ \frac{\partial^2 l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \right]_{\boldsymbol{\beta}=\boldsymbol{\beta}^{(0)}} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^{(0)}) = \mathbf{0}$$

Entonces, sea  $J(\boldsymbol{\beta}) = -\frac{\partial^2 l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T}$ , también llamada la matriz de información observada de Fisher, haciendo un desarrollo adicional sobre la expresión anterior se tiene que

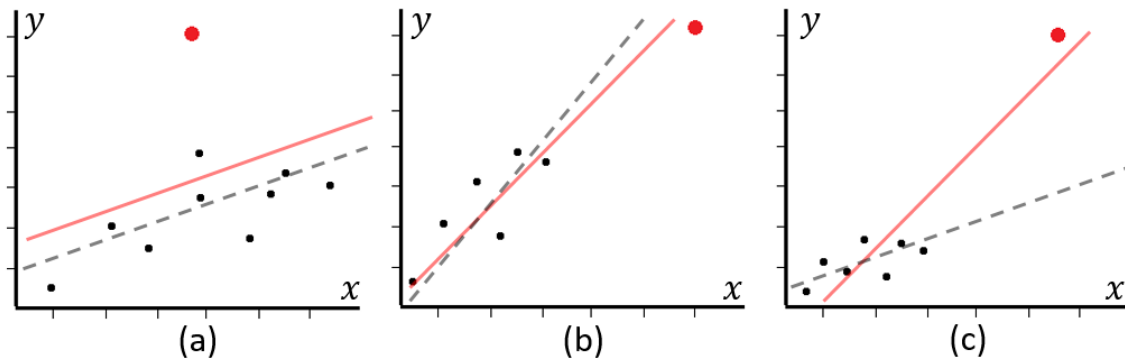
$$\hat{\boldsymbol{\beta}} \approx \boldsymbol{\beta}^{(0)} + [J(\boldsymbol{\beta}^{(0)})]^{-1} U(\boldsymbol{\beta}^{(0)})$$

De esta expresión se derivan métodos numéricos como el Newton-Raphson y Scoring de Fisher con los cuales se obtiene el estimador máximo verosimil del MLG descrito. A continuación se describe brevemente el algoritmo Newton-Raphson:

- (0) Fijar  $m = 0$ ,  $\xi > 0$  y establecer el valor inicial  $\boldsymbol{\beta}^{(0)}$ .
- (1) Calcular  $\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} + [J(\boldsymbol{\beta}^{(m)})]^{-1} U(\boldsymbol{\beta}^{(m)})$
- (2) Calcular  $\delta(m+1) = \max \left\{ \frac{|\beta_1^{(m+1)} - \beta_1^{(m)}|}{|\beta_1^{(m)}|}, \dots, \frac{|\beta_p^{(m+1)} - \beta_p^{(m)}|}{|\beta_p^{(m)}|} \right\}$ .
- (3) Fijar  $m = m + 1$ .
- (4) Repetir pasos (1)-(3) hasta que  $\delta(m) < \xi$ .
- (5)  $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}^{(m)}$ .

## 2.3. Análisis de Influencia

Antes de entrar en detalle acerca de las medidas y métodos propuestos para la identificación de observaciones influyentes en MLG, consideremos los casos de la Figura (2-1) presentados en McCullagh and Nelder (1989) para un modelo de regresión lineal simple:



**Figura 2-1.:** Puntos outlier, de alto leverage e influyentes. Tomado de McCullagh and Nelder (1989).

En los gráficos de la Figura (2-1) en color rojo se marca la observación de interés de cada caso. La línea continua de color rojo corresponde a la recta de regresión estimada con todas las observaciones y la línea punteada de color gris es la estimación de la recta sin considerar la observación resaltada en rojo. En la Figura (a) se marca una observación que presenta un valor atípico en la variable respuesta  $y$ . Puede observarse que al omitirse dicha observación, cambia la estimación del intercepto de la recta, pero aparentemente no se altera la estimación de la pendiente. En el Figura (b) se resalta una observación que presenta un valor atípico tanto en la respuesta  $y$  como en la covariable  $x$ . Sin embargo, esta observación es consistente con el modelo, por lo cual no afecta de forma drástica la estimación de la recta de regresión. Por último, en la Figura (c) se muestra una observación alejada en la respuesta  $y$  y en la covariable. Contrario a lo que ocurre en el caso de la Figura (b), la observación señalada es influyente en la estimación de la pendiente y el intercepto del modelo, puesto que no es consistente con la recta de regresión obtenida al omitir esta observación.

Siguiendo las definiciones de los autores, se dice que un *outlier* corresponde a una observación que presenta un valor atípico en la variable respuesta. Lo cual caracteriza a las observaciones marcadas en los tres casos. Por otra parte, una observación cuyas covariables se encuentran alejadas del espacio de las mismas, se denomina un punto de alto *leverage*, lo cual es una característica de los puntos rojos de los casos (b) y (c). Ahora, un *outlier* o un punto de alto *leverage* pueden ser influyentes o no dependiendo de la relación dada entre la respuesta y las covariables. Del ejemplo anterior, se tiene que en el caso (b) aunque la observación marcada es un *outlier* y un punto de alto *leverage*, resulta no ser influyente en la estimación de los

parámetros, contrario al punto del caso (c).

Como se mencionó anteriormente, la medida más usada para determinar si una observación es influyente o no es la distancia de Cook de la expresión (1-3), propuesta por Cook (1977). A partir de esta medida se han propuesto algunas otras que son modificaciones de la propuesta por Cook. Chien and Tsou (2014), por ejemplo, proponen la siguiente medida para modelos de respuesta continua, basada en una estimación robusta a desviaciones de la distribución asociada al modelo,

$$AW_i = \left( \hat{\beta} - \hat{\beta}_{(i)} \right)^T \widehat{Var}_{adj}^{-1}(\hat{\beta}_{(i)}) \left( \hat{\beta} - \hat{\beta}_{(i)} \right) ,$$

donde  $\widehat{Var}_{adj}(\hat{\beta})$  es la estimación de la matriz de covarianzas de  $\hat{\beta}$ , obtenida a partir de la estimación robusta a desviaciones del supuesto de normalidad en un modelo de regresión. Por otra parte, Jang and Anderson-Cook (2015) proponen un método gráfico denominado *Firework plot* como el de la Figura (2-2), a través del cual se identifican las observaciones influyentes observando las que presentan cambios mayores tanto en las estimaciones de los parámetros, como en el deviance asociado al modelo ajustado.

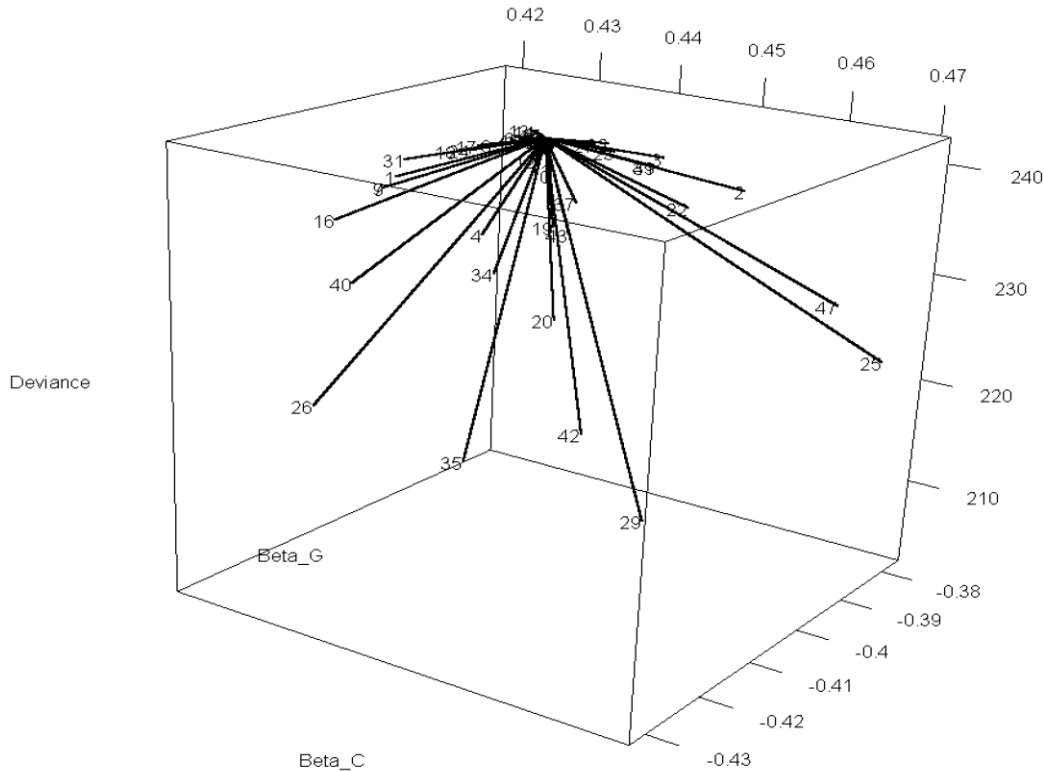


Figura 2-2.: Firework plot. Tomado de Jang and Anderson-Cook (2015)

### 3. Análisis de Datos Funcionales

Como se mencionó anteriormente, en el Análisis de Datos Funcionales se estudian objetos que puede ser interpretados como funciones. Luego, dada la naturaleza de este tipo de objetos, se dice que estos habitan en un espacio infinito dimensional puesto que, en teoría, cada función podría observarse en un número infinito de puntos dentro de su dominio. El espacio usual sobre el que se trabajan los datos funcionales es el espacio de funciones cuadrado integrables,  $L^2(\Omega)$ . Este espacio tiene propiedades interesantes puesto que es un espacio de Hilbert. Es decir, que corresponde a un espacio vectorial con producto interno que es completo. Razón por la cual gran parte de la teoría se ha desarrollado sobre él, dado que su estructura matemática tiene similitudes a la de  $\mathbb{R}^p$ , lo que permite extender conceptos de la estadística multivariada a los datos funcionales de forma intuitiva [Happ (2020)]. A continuación se precisan algunos conceptos sobre este tipo de espacios vectoriales:

**Definición 1.** Una sucesión de elementos  $\{X_n\}$  en un espacio métrico  $(M, d)$  se llama sucesión de Cauchy si para todo  $\epsilon > 0$  existe un entero  $N$  tal que  $d(X_n, X_m) < \epsilon$  si  $n \geq N$  y  $m \geq N$ .

**Definición 2.** Un espacio métrico  $(M, d)$  es completo si toda sucesión de Cauchy converge a un elemento de  $M$ .

$$\lim_{n \rightarrow \infty} X_n = X$$

donde  $X \in M$ .

**Definición 3.** Un espacio de Hilbert es un espacio vectorial con producto interno que es completo.

Ahora, a modo de ilustración de los conceptos, por ahora consideraremos funciones de una variable. Sin embargo, los siguientes resultados se pueden generalizar de forma directa a funciones de más variables haciendo uso de integrables dobles, triples, etc.

El espacio de  $L^2$  se define como  $L^2 := \{f : \int f^2(t)dt < \infty\}$ . Sea  $X$  un elemento aleatorio de dicho espacio, se dice que  $X$  es cuadrado integrable si  $E[||X||^2] < \infty$ . La norma  $|| \cdot ||$  es la norma inducida por el producto punto  $\langle f, g \rangle = \int f(t)g(t)dt$ . Luego para una función  $f$ , su norma está dada por  $||f|| = \langle f, f \rangle^{1/2}$ . Así,  $E[||X||^2] = E(\int X^2(t)dt) < \infty$ .

En el caso de los datos funcionales, se define el operador de covarianza, el cual es análogo a la matriz de covarianzas de un proceso multivariado. Este operador resulta ser una descripción

adecuada de la estructura de segundo orden de un espacio de Banach en general [Kokoszka and Reimherr (2017)]. Sea  $X$  un elemento aleatorio de  $L^2$  tal que  $E[X] = \mu$ . Se define el operador de covarianza asociado al proceso estocástico  $X$ , evaluado en una función  $y \in L^2$ , en un punto  $t$ ,  $\Gamma_X(y)(t)$

$$\begin{aligned} \Gamma_X(y)(t) &:= E[\langle X - \mu, y \rangle (X - \mu)] \\ &= E \left[ \int (X - \mu)(s)y(s)ds (X - \mu)(t) \right] = \int E[(X - \mu)(s)(X - \mu)(t)] y(s)ds \\ &= \int Cov[X(s), X(t)] y(s)ds = \int Cov(t, s)y(s)ds \end{aligned} \quad (3-1)$$

Puede observarse que este depende de la covarianza del procesos en dos puntos de su dominio  $s$  y  $t$ . La expresión  $\langle X - \mu, y \rangle (X - \mu)$  corresponde al producto tensor para un espacio de Hilbert. Para el caso multivariado, el producto tensor análogo es  $(X - \mu) \otimes (X - \mu)$ , a partir del cual se construye la matriz de covarianzas:

$$\Sigma = E \left[ (X - \mu) \otimes (X - \mu) \right] = E \left[ (X - \mu)(X - \mu)^T \right].$$

Del operador  $\Gamma_X$  se deducen algunos resultados de los cuales se derivan varios de los métodos frecuentes en el análisis de datos funcionales, como el análisis de componentes principales funcionales o FPCA, por sus siglas en inglés. En este caso, los componentes principales corresponden a funciones propias que satisfacen la siguiente relación, donde  $v_j(t)$  es la función propia asociada al valor propio  $\lambda_j$ , sobre la cual se aplica el operador de la expresión (3-1).

$$\Gamma_X(v_j)(t) = \lambda_j v_j(t).$$

### 3.1. Modelos con datos funcionales

Además del FPCA otros análisis clásicos de la estadística se han adaptado para datos funcionales. Ese es el caso de los modelos lineales, sobre los cuales se han propuesto una amplia gama de modelos con respuesta escalar o funcional y cuyas covariables pueden ser escalares, funcionales o combinación de las dos [Ramsay and Silverman (2005)]. Suponga, por ejemplo, que se desea explicar la precipitación anual promedio a partir de perfiles de temperatura medidos en un año particular. Entonces, un modelo adecuado podría ser el siguiente

$$y_i = \alpha + \int_T \beta(s)X_i(s)ds + e_i, \quad (3-2)$$

donde  $y_i$  es el promedio de precipitación anual de una ciudad  $i$ ,  $X_i(t)$  es el perfil de temperatura de la ciudad  $i$  durante el año,  $e_i$  es la componente aleatoria de error de media 0



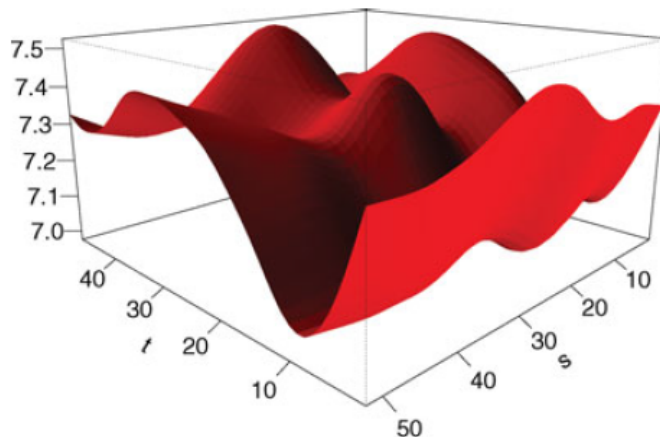
y  $\beta(s)$  es la función a estimar, la cual mide la relación entre el perfil de temperatura y la precipitación promedio anual, en todo el intervalo de tiempo de observación. Esta ecuación corresponde a la de un modelo lineal con respuesta escalar y covariable funcional. Si por otra parte, se midiera el perfil de precipitación de la misma ciudad a lo largo del año y se deseara explicar la relación entre dicho perfil con el de temperatura, se puede considerar un modelo con respuesta funcional como el siguiente:

$$y_i(t) = \alpha + \int_T \beta(t, s) X_i(s) ds + e_i(t), \quad (3-3)$$

donde  $y_i(t)$  es el perfil de precipitación de una ciudad  $i$ ,  $X_i(t)$  es el perfil de temperatura de la ciudad  $i$  durante el año,  $e_i(t)$  es la función aleatoria de error y  $\beta(t, s)$  es la superficie a estimar, la cual mide la relación entre el perfil de temperatura y el de precipitación. Cabe resaltar que los modelos anteriores se pueden extender de forma relativamente directa para imágenes (o superficies). Por ejemplo, [Ivanescu \(2018\)](#), propone el siguiente modelo donde la respuesta es una superficie, al igual que la covariable, para explicar la relación entre el perfil de la demanda de energía medida en 48 tiempos durante el día por 52 semanas del año, en función del perfil de temperatura medido en los mismos intervalos de tiempo.

$$Y_i(t_{ij}, s_{ik}) = \alpha(t_{ij}, s_{ik}) + \beta(t_{ij}, s_{ik}) X_i(t_{ij}, s_{ik}) + e_i(t_{ij}, s_{ik}). \quad (3-4)$$

En el modelo anterior,  $Y_i(t_{ij}, s_{ik})$  corresponde a la demanda de energía de la  $i$ -ésima unidad muestral medida en el tiempo del día  $t_{ij}$ ,  $j = 1, \dots, 48$ , y en la semana del año  $s_{ik}$ ,  $k = 1, \dots, 52$ . Análogamente,  $X_i(t_{ij}, s_{ik})$  corresponde a la medición de la temperatura. En este ejemplo, el parámetro a estimar  $\beta(t_{ij}, s_{ik})$  corresponde a una superficie como la de la siguiente Figura.



**Figura 3-1.:** Ejemplo de estimación del parámetro superficie  $\beta$ . Tomado de [Ivanescu \(2018\)](#).

Por último, el interés de este trabajo se centra en los modelos lineales generalizados con respuesta escalar y con covariables escalares e imágenes. Específicamente, se estudiarán los modelos descritos por la expresión (1-1). En la literatura existen algunas propuestas en este campo, las cuales se basan, por ejemplo, en regresión tensor, regresión lasso y métodos basados en componentes principales funcionales (FPCA). Por otra parte, la propuesta de Wang et al. (2017) supone que el parámetro imagen del modelo es una función suave a trozos que puede presentar saltos y bordes desconocidos, lo cual es frecuente al trabajar con imágenes reales. Bajo dicho supuesto, espacios como  $L^2$  o subconjuntos de este, como los espacios de Sobolev, resultan no ser adecuados pues están restringidos a funciones diferenciables. En ese sentido, los autores suponen que el parámetro imagen  $\beta$  pertenece al espacio de variación acotada  $BV(\Omega)$ , el cual es un espacio que contiene funciones continuas a trozos con saltos y bordes.

En general, se dice que una función  $\beta_0$  es una función de variación acotada si la variación total de  $\beta_0$ ,  $\|\beta_0\|_{TV}$ , es finita. Es decir, si la siguiente condición se satisface

$$\|\beta_0\|_{TV} := \sup \left\{ \int_{\Omega} \beta_0(u, v) \operatorname{div} f(u, v) du dv \right\} < \infty,$$

siendo  $f(u, v) = (f_1(u, v), f_2(u, v))$  tal que  $f \in C_c^\infty(\Omega, \mathbb{R}^2)$  y  $|f|_\infty = \sup_{(u,v) \in \Omega} |f(u, v)| \leq 1$  y  $\operatorname{div} f(u, v) = \frac{\partial}{\partial u} f_1(u, v) + \frac{\partial}{\partial v} f_2(u, v)$ , donde  $C_c^\infty(\Omega, \mathbb{R}^2)$  corresponde a un campo vectorial con valor en  $\mathbb{R}^2$ .

## 3.2. MLG escalar-función estimado via variación total

A continuación se expondrán los detalles principales del Modelos Lineal Generalizado Escalar-Imagen estimado vía variación total, o GSIRM-TV por sus siglas en inglés (*Generalized Scalar-on-Image Regression Models*), propuesto por Wang et al. (2017). A través de los modelos GSIRM-TV se consideran un conjunto de predictores  $Z \in \mathbb{R}^p$  y  $X \in \mathbb{R}^{N \times N}$  (una imagen) observados sin error. Se asume que

$$\begin{aligned} Y|(X, Z) &\sim \text{Familia Exponencial Lineal } (\mu, \phi) \\ g(\mu) &= \theta^t Z + \langle X, \beta \rangle \end{aligned} \tag{3-5}$$

Como se mencionó anteriormente,  $X$  y  $\beta$  son imágenes que pertenecen al espacio infinito-dimensional  $BV(\Omega)$ . Sin embargo, dado que las imágenes solo es posible observarlas en un conjunto finito de puntos (píxeles), el producto punto de la expresión anterior se define en términos de dicho conjunto finito. Por lo cual en adelante se asume que  $\langle U, V \rangle = \sum_{i,j} u_{i,j} v_{i,j}$  para  $U = (u_{i,j}), V = (v_{i,j}) \in \mathbb{R}^{N \times N}$ .

### Estimación del modelo de regresión normal

A modo de ilustración, considere el siguiente modelo de regresión lineal:

$$Y = \langle X, \beta \rangle + \varepsilon \quad (3-6)$$

con  $E(\varepsilon|X) = 0$ ,  $E(\varepsilon^2|X) = \sigma^2$  y  $E(X) = 0$ ,  $E(Y) = 0$ . Sea  $\Omega = [0, 1]^2$ . Si el parámetro imagen  $\beta$ , perteneciera al espacio de funciones cuadrado integrables. Esto es, si  $\beta_0 \in L^2(\Omega)$ , el modelo podría escribirse como

$$Y = \int_0^1 \int_0^1 x(u, v) \beta(u, v) dudv + \varepsilon,$$

expresión similar al modelo con respuesta escalar y covariable funcional con dominio univariado descrita en la ecuación (3-2). Sin embargo, se asume que  $\beta$  es una función del espacio  $BV(\Omega)$ .

Considerando el modelo (3-6) y el espacio  $BV(\Omega)$ , para la estimación de  $\beta_0$ , [Wang et al. \(2017\)](#) proponen minimizar la siguiente expresión

$$\|\beta\|_{TV} \text{ sujeto a } \sum_{i=1}^n (Y_i - \langle X_i, \beta \rangle)^2 \leq \lambda^2, \quad (3-7)$$

donde  $\lambda$  es un hiperparámetro que controla el nivel de ruido, o equivalentemente,

$$\sum_{i=1}^n (Y_i - \langle X_i, \beta \rangle)^2 + \check{\lambda} \|\beta\|_{TV}, \quad (3-8)$$

siendo  $\check{\lambda}$  un hiperparámetro diferente a  $\lambda$ , que cumple la misma función. Para minimizar (3-7) o (3-8), abusando de la notación, se supone que  $\beta \in \mathbb{R}^{N \times N}$ , es decir, se supone que el parámetro imagen es un bloque de pixeles de dimensión  $\mathbb{N} \times \mathbb{N}$ . Dado que el espacio  $BV(\Omega)$  es un espacio infinito-dimensional, es necesario definir la variación total en su versión discreta. Por lo tanto, para  $\beta \in BV(\Omega)$ , se define el gradiente discreto  $\nabla : BV(\Omega) \rightarrow \mathbb{R}^{N \times N \times 2}$  de la siguiente manera:

$$(\nabla \beta)_{jk} = \begin{cases} (\beta_{j+1,k} - \beta_{jk}, \beta_{j,k+1} - \beta_{jk}) & \text{si } 1 \leq j, k \leq N-1, \\ (0, \beta_{j,k+1} - \beta_{jk}) & \text{si } j = N, 1 \leq k \leq N-1, \\ (\beta_{j+1,k} - \beta_{jk}, 0) & \text{si } k = N, 1 \leq j \leq N-1, \\ (0, 0) & \text{si } k = j = N. \end{cases}$$

Luego, se define la variación total (discreta) en su versión anisotrópica

$$\|\beta\|_{TV}^{aniso} = \|\nabla \beta\|_1 = \sum_{j,k} \{ |(\nabla \beta)_{jk,1}| + |(\nabla \beta)_{jk,2}| \}. \quad (3-9)$$

La versión isotrópica de la norma se define como

$$\|\beta\|_{TV}^{iso} = \sum_{jk} \|(\nabla\beta)_{jk}\|_2 = \sum_{j,k} \sqrt{(\nabla\beta)_{jk,1}^2 + (\nabla\beta)_{jk,2}^2}, \quad (3-10)$$

aunque ambas normas son equivalentes, en el sentido de equivalencia de dos normas definidas sobre un espacio vectorial.

A modo de ilustración, en la Figura (3-2) se muestra una imagen observada (izquierda) y las dos componentes del gradiente discreto asociadas (centro y derecha). Se observa que las componentes del gradiente discreto poseen la información de cómo varía la imagen pixel a pixel en sentido vertical (centro) y horizontal (derecha).



**Figura 3-2.:** Imagen observada (izquierda) y componentes del gradiente discreto: sentido vertical (centro) y sentido horizontal (derecha). Tomado de Wang et al. (2017).

Ahora, sea  $A_X$  la matriz de diseño  $n \times N^2$  cuya fila  $i$ -ésima es  $vec(X_i)$ , escribiendo (3-7) matricialmente,

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \|\beta\|_{TV} \quad \text{sujeto a} \quad \|Y - A_X\beta\|_2 \leq \lambda.$$

La optimización de la variación total es un procedimiento usado frecuentemente para la reconstrucción de imágenes. Para resolver el problema descrito en la expresión anterior, se emplea el algoritmo TVAL3 basado en el método lagrangiano aumentado. Así, se resuelve el problema de optimización equivalente:

$$\underset{\omega, \beta}{\operatorname{min}} \sum_{l=1}^{N^2} \|\omega_l\|_1 \quad \text{sujeto a} \quad \|Y - A_X\beta\|_2 \leq \lambda,$$

donde  $\omega_l = D_l\beta$ , siendo  $D_l$  una matriz de constantes  $2 \times N^2$  asociadas con el gradiente discreto. Siguiendo el ejemplo Wang et al. (2017), supongamos que las imágenes observadas son de dimensiones  $2 \times 2$  ( $N = 2$ ), entonces la imagen parámetro vectorizada es

$$\beta = (\beta_{11}, \beta_{21}, \beta_{12}, \beta_{22}). \quad \text{Se pueden escoger } D_1 = \begin{pmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 1 \end{pmatrix}, \quad D_2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 \end{pmatrix},$$

$$D_3 = \begin{pmatrix} 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad D_4 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \text{de modo que } D_1\beta = (\nabla\beta)_{11}, \quad D_2\beta = (\nabla\beta)_{12}, \\ D_3\beta = (\nabla\beta)_{21} \quad \text{y} \quad D_4\beta = (\nabla\beta)_{22}.$$

La correspondiente función lagrangiana aumentada está dada por:

$$L_A(\omega, \beta) = \sum_{l=1}^{N^2} \|\omega_l\|_1 - v_l^T (D_l \beta - \omega_l) + \frac{\alpha_l}{2} \|D_l \beta - \omega_l\|_2^2 + \frac{\gamma}{2} \|A_X \beta - Y\|_2^2.$$

El elemento que minimiza el problema descrito puede obtenerse iterativamente a través de algoritmo TVAL3 que se describe a continuación.

1. Fijar valores iniciales  $\beta^{(0)}, v_l^{(0)}$ .
2. Dados  $\beta = \beta^{(k)}$  y  $v_l = v_l^{(k)}$ , encontrar  $\omega^{(k+1)}, l = 1, \dots, N^2$ , minimizando

$$\|\omega_l\|_1 - v_l^T (D_l \beta - \omega_l) + \frac{\alpha_l}{2} \|D_l \beta - \omega_l\|_2^2$$

3. Dados  $\omega = \omega^{(k)}$  y  $v = v_l^{(k)}, l = 1, \dots, N^2$ , minimizando

$$\sum_{l=1}^{N^2} \left\{ -v_l D_l \beta + \frac{\alpha_l}{2} \|D_l \beta - \omega_l\|_2^2 + \frac{\gamma}{2} \|A_X \beta - Y\|_2^2 \right\}$$

4. Actualizar  $v_l^{(k+1)}$  con

$$v_l^{(k+1)} = v_l^{(k)} - \alpha_l (D_l \beta^{(k+1)} - \omega_l^{(k+1)})$$

5. Iterar 2-4 hasta convergencia.

### Estimación del modelo de lineal generalizado

Para el caso del modelo lineal generalizado escalar-imagen, considere una covariable imagen  $X \in \mathbb{R}^{N \times N}$  y un vector de covariables escalares  $Z \in \mathbb{R}^p$ . Se asume que  $Y$  tiene una distribución que pertenece a la familia exponencial. Es decir, su función de densidad se puede expresar como

$\exp(\{Y\eta(X, Z; \theta, \beta) - b(\eta(X, Z; \theta, \beta))\}/a(\psi) + c(y, \psi))$ , expresión equivalente a la dada en la ecuación (2-1), donde  $a(\cdot), b(\cdot)$  y  $c(\cdot)$  son funciones conocidas al igual que valor  $\psi$ .

Se puede mostrar que

$$E(Y|X) = \mu(X, Z; \theta_0, \beta_0) = b'(\eta(X, Z; \theta_0, \beta_0))$$

y

$$Var(Y|X) = a(\psi)b''(\eta(X, Z; \theta_0, \beta_0))$$

el cual es un resultado de la teoría de los modelos lineales generalizados [McCullagh and Nelder (1989)]. Para la distribución normal con varianza  $\sigma^2$ , por ejemplo,  $a(\psi) = \sigma^2$  y  $b(\eta) = \eta^2/2$ . Para una distribución Bernoulli,  $a(\psi) = 1$  y  $b(\eta) = \log(1 + e^\eta)$ .

Sea  $\xi = (\theta, \beta) \in R^p \times BV(\Omega)$ . Dados los datos observados, se propone maximizar la función de verosimilitud penalizada dada por

$$\frac{1}{n} \sum_{i=1}^n \{Y_i \eta(X_i, Z_i; \theta, \beta) - b(\eta(X_i, Z_i; \theta, \beta))\} + \lambda \|\beta\|_{TV} . \quad (3-11)$$

Para el cálculo de  $\hat{\xi} = (\hat{\theta}, \hat{\beta})$ , se modifica el procedimiento iterativo estándar de mínimos cuadrados ponderados para MLG, añadiendo el término de penalización de variación total. Como en el caso del modelo lineal clásico, la estimación se hace de forma iterativa. Sea  $\hat{\xi}_I$  una estimación de prueba de  $\xi$ , se introducen pesos y variables dependientes iterativos

$$\hat{\omega}_{i,I} = \ddot{b}(\hat{\eta}_{i,I}) \quad \text{y} \quad \hat{Y}_{i,I} = g(\hat{\mu}_{i,I}) + (Y_i - \hat{\mu}_{i,I}) \dot{g}(\hat{\mu}_{i,I}) . \quad (3-12)$$

Luego, se calcula una nueva estimación de  $\xi$  ( $\hat{\xi}_{I+1}$ ) minimizando

$$\sum_{i=1}^n \hat{\omega}_{i,I} \left( \hat{Y}_{i,I} - \frac{\partial}{\partial \xi} \mu_{i,I}(\hat{\xi}_I) \xi \right)^2 + \lambda \|\beta\|_{TV} , \quad (3-13)$$

o la expresión matricial equivalente

$$\|W_I^{1/2}(\hat{Y}_I - \mu'_I \xi)\|_2^2 + \lambda \|\beta\|_{TV} \quad (3-14)$$

con

$$W_I^{1/2} = \begin{pmatrix} \sqrt{\hat{\omega}_{1,I}} & 0 & \cdots & 0 \\ 0 & \sqrt{\hat{\omega}_{2,I}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\hat{\omega}_{n,I}} \end{pmatrix}, \quad \mu'_I = \begin{pmatrix} \frac{\partial \mu_{1,I}}{\partial \theta} & \frac{\partial \mu_{1,I}}{\partial \beta} \\ \frac{\partial \mu_{2,I}}{\partial \theta} & \frac{\partial \mu_{2,I}}{\partial \beta} \\ \vdots & \vdots \\ \frac{\partial \mu_{n,I}}{\partial \theta} & \frac{\partial \mu_{n,I}}{\partial \beta} \end{pmatrix}, \quad \hat{Y}_I = \begin{pmatrix} \hat{Y}_{1,I} \\ \hat{Y}_{2,I} \\ \vdots \\ \hat{Y}_{n,I} \end{pmatrix} .$$

Este problema de optimización puede ser resuelto con el algoritmo TVAL3 y de esta forma se obtiene la estimación del parámetro imagen  $\beta_0$  y del parámetro vector  $\theta$ .

**Ejemplo 2.** Considere el modelo logístico  $g(\mu_i) = \langle X_i, \beta_0 \rangle + \theta_0^t Z_i$ , con  $g(\mu_i) = \ln(\mu_i/(1-\mu_i))$ , dado  $\hat{\xi}_I$ , se pueden obtener los pesos y respuestas efectivas teniendo en cuenta lo siguiente:

$$\begin{aligned} \eta &= \ln\left(\frac{\mu}{1-\mu}\right) \\ e^\eta &= \frac{\mu}{1-\mu} \\ e^\eta - \mu e^\eta &= \mu \\ \mu &= \frac{e^\eta}{(1+e^\eta)} \end{aligned}$$

$y$

$$\begin{aligned} b(\eta) &= -\ln(1 - \mu) = -\ln\left(1 - \frac{e^\eta}{1 + e^\eta}\right) \\ &= -\ln\left(\frac{1 + e^\eta - e^\eta}{1 + e^\eta}\right) = -\ln\left(\frac{1}{1 + e^\eta}\right) = \ln(1 + e^\eta) \end{aligned}$$

Luego,

$$\begin{aligned} b(\eta) &= \ln(1 + e^\eta) \\ \dot{b}(\eta) &= \frac{e^\eta}{1 + e^\eta} \\ \omega &= \ddot{b}(\eta) = \frac{(1 + e^\eta)e^\eta - e^\eta e^\eta}{(1 + e^\eta)^2} = \frac{e^\eta}{(1 + e^\eta)^2} \end{aligned}$$

Si tomamos  $g(\mu) = \eta = \ln\left(\frac{\mu}{1-\mu}\right)$ ,

$$\begin{aligned} \dot{g}(\mu) &= \frac{(1 - \mu + \mu)/(1 - \mu)^2}{\mu/(1 - \mu)} = \frac{1}{\mu(1 - \mu)} \\ \hat{Y} &= g(\mu) + (Y - \mu)\dot{g}(\mu) = \eta + \frac{(Y - \mu)}{\mu(1 - \mu)} \end{aligned}$$

Así, se muestra que

$$\hat{\omega}_{i,I} = \frac{e^{\hat{\eta}_{i,I}}}{(1 + e^{\hat{\eta}_{i,I}})^2} \quad y \quad \hat{Y}_{i,I} = \hat{\eta}_{i,I} + \frac{Y_i - \hat{\mu}_{i,I}}{\hat{\mu}_{i,I}(1 - \hat{\mu}_{i,I})}$$

### 3.3. Medidas de influencia en modelos escalar-función

Para identificar observaciones que afecten de forma importante la estimación de parámetros en modelos con respuesta escalar y cuya covariable es una imagen, como por ejemplo el modelo GSIRM-TV descrito en la expresión (3-5), actualmente no existen métodos propuestos en la literatura. Sin embargo, [Febrero-Bande et al. \(2010\)](#) proponen dos medidas basadas en la distancia de Cook bajo el modelo de respuesta escalar descrito en (3-2), donde la covariable funcional es de dominio univariado y cuyo método de estimación se basa en análisis de componentes principales funcionales. Considerando el modelo (3-2), los autores proponen la siguiente medida de Cook funcional, la cual se interpreta como el cambio estandarizado en el vector de predicciones después de eliminar la observación  $i$ -ésima:

$$CP_i = \frac{(\hat{\mathbf{y}}_i - \hat{\mathbf{y}}_{(-i)})^t (\hat{\mathbf{y}}_i - \hat{\mathbf{y}}_{(-i)})}{s_R^2}, \quad (3-15)$$

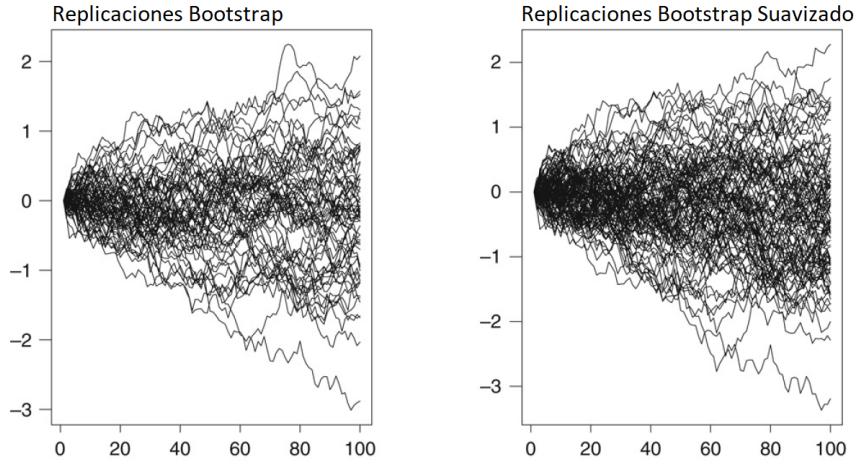
donde  $\hat{\mathbf{y}}$  y  $\hat{\mathbf{y}}_{(-i)}$  son los vectores con las predicciones obtenidas a partir del modelo estimado con todas las observaciones y tras eliminar la observación  $i$ -ésima, respectivamente, y  $s_R^2$  es el estimador de la varianza de la componente de error  $e_i$ . La segunda medida de Cook funcional propuesta por los autores corresponde a una medida estandarizada del cambio en el parámetro funcional al eliminar la observación  $i$ -ésima, y está dada por la siguiente expresión:

$$CE_i = \frac{\|\hat{\beta} - \hat{\beta}_{(-i)}\|^2}{\frac{s_R^2}{n} \sum_{k=1}^{k_n} 1/\lambda_k}, \quad (3-16)$$

donde  $\hat{\beta}$  y  $\hat{\beta}_{(-i)}$  son las estimaciones del parámetro funcional  $\beta$  obtenidas a partir del modelo estimado con todas las observaciones y omitiendo la observación  $i$ -ésima, respectivamente.  $k_n$  corresponde al número de componentes principales escogido y  $\lambda_k$ ,  $k = 1, \dots, k_n$  los valores propios asociados. Dada la dificultad de encontrar una distribución aproximada para las medidas de influencia, aún considerando procesos gaussianos, los autores proponen un método basado en bootstrap, denominado “bootstrap suavizado”, para aproximar la distribución de las medidas de influencia y determinar el valor de referencia a partir del cual una observación es considerada influyente. A continuación se describe en detalle este método:

1. Ajustar el modelo lineal a los datos observados  $(X_1, y_1), \dots, (X_n, y_n)$  y obtener las medidas de influencia  $IF_1, \dots, IF_n$ , siendo  $IF_i = CP_i$  o  $IF_i = CE_i$ ,  $i = 1, \dots, n$ .
2. Obtener  $B$  muestras bootstrap de tamaño  $n$  de los residuos  $e_1, \dots, e_n$  y  $B$  muestras bootstrap de tamaño  $n$  de las covariables funcionales  $X_1, \dots, X_n$ .
3. Suavizar las muestras bootstrap de la siguiente forma:  
 Calcular  $\tilde{X}_i^b = X_i^b + Z_i^b$  donde  $Z_i^b$  es un proceso gaussiano con operador de covarianza  $\gamma_x \Gamma_x$ , donde  $\gamma_x$  es un hiperparámetro predeterminado.  
 Calcular  $\tilde{e}_i^b = e_i^b + z_i^b$  donde  $z_i^b \sim N(0, \gamma_e S_R^2)$ ,  $i = 1, \dots, B$ , siendo  $\gamma_e$  un hiperparámetro predeterminado.
4. Calcular  $\tilde{y}_i^b = \int_T \tilde{X}_i^b(t) \hat{\beta}_{(k_n)}(t) dt + \tilde{e}_i^b$ .
5. Para cada muestra bootstrap suavizada obtener las medidas de influencia  $\tilde{IF}_1^b, \dots, \tilde{IF}_n^b$ .





**Figura 3-3.:** Replicaciones bootstrap de un conjunto de curvas (Izquierda). Replicaciones por el método bootstrap suavizado de un conjunto de curvas (Derecha). Tomado de [Febrero-Bande et al. \(2010\)](#)

El cuantil estimado para cada medida de influencia  $IF_i$ ,  $i = 1, \dots, n$ , es

$$p_i^B = \frac{\sum_{b=1}^B \sum_{h=1}^n I \{ \tilde{IF}_h^b \leq IF_i \}}{nB}$$

Luego, se dice que una observación es influyente, si el cuantil asociado es, por ejemplo, mayor a 0.95.

Ahora, respecto al procedimiento bootstrap suavizado, los autores señalan que al utilizar este método en lugar del bootstrap clásico, se evita la repetición de covariables (en este caso curvas) en cada muestra bootstrap. Por el contrario, se obtienen muestras de covariables que conservan las características de las covariables originales y que no se repiten dentro de la muestra como se ilustra en la Figura (3-3).

### 3.4. Detección de outliers en conjuntos de imágenes

Como se mencionó anteriormente, en la actualidad no existen propuestas en la literatura para la identificación de observaciones influyentes en modelos escalar-imagen, en particular en los modelos GSIRM-TV. Una alternativa podría ser adaptar el procedimiento bootstrap suavizado propuesto por [Febrero-Bande et al. \(2010\)](#), descrito en la sección anterior, definiendo el método de suavizamiento de las imágenes  $X_i$  y las covariables escalares  $Z_i$ , similar al propuesto por los autores, así como las medidas de influencia apropiadas. Es decir, adicionando la realización de un proceso estocástico de un espacio vectorial adecuado, en cada observación dentro de cada muestra bootstrap. Esto es, definir la forma de calcular  $\tilde{X}_i^b = X_i^b + V_i^b$  y  $\tilde{Z}_i^b = Z_i^b + U_i^b$ , donde  $V_i^b$  es la realización de un proceso estocástico definido en el espacio de

variación acotada  $BV(\Omega)$ , y  $U_i^b$  la realización de un proceso estocástico definido en  $\mathbb{R}^p$ , por ejemplo. Sin embargo, llevar a cabo esta propuesta o cualquier método basado en bootstrap, requiere un alto costo computacional puesto que, bajo un tamaño de muestra  $n$  y considerando  $B$  muestras bootstrap, se requiere estimar los parámetros del modelo  $n(B + 1) + 1$  veces. Lo cual se traduce en un costo computacional bastante alto, dadas las características del algoritmo TVAL3 para estimar los parámetros de un modelo GSIRM-TV, incluso cuando las imágenes son de baja resolución.

Ahora, en teoría, si una observación es influyente, generará una afectación evidente en la estimación de los parámetros del modelo. Al omitir dicha observación, se espera que la estimación de los parámetros cambie considerablemente. Por lo tanto, una forma de identificar un cambio grande y en consecuencia la observación influyente es comparando la estimación obtenida al omitir dicha observación, con el conjunto de estimaciones resultado de omitir cada una de las observaciones restantes. Así, a través de este trabajo se propone identificar observaciones influyentes en la estimación del parámetro imagen, identificando estimaciones alejadas del conjunto de  $n$  estimaciones obtenidas omitiendo cada una de las unidades muestrales. Para esto, en este trabajo se propone utilizar los métodos basados en *outlyingness* propuestos por [Rousseeuw and Hubert \(2018\)](#) y [Dai and Genton \(2018\)](#), para la identificación de observaciones *outliers* en conjuntos de datos funcionales, en particular de imágenes. En decir, en el contexto de los modelos GSIRM-TV, se propone identificar observaciones influyentes aplicando los métodos en mención sobre el conjunto de estimaciones  $\{\hat{\beta}_{(1)}, \dots, \hat{\beta}_{(n)}\}$ , donde  $\hat{\beta}_{(i)}$  es la estimación del parámetro imagen  $\beta$  obtenida al omitir la observación  $i$ -ésima de la muestra. De esta forma, se puede definir un criterio claro para determinar si una estimación está alejada del conjunto de estimaciones, evitando el costo computacional de un procedimiento basado en remuestreo.

Teniendo en cuenta lo anterior, es importante entender el concepto de *outlyingness* usado en las propuestas de [Rousseeuw and Hubert \(2018\)](#) y [Dai and Genton \(2018\)](#). Este concepto se relaciona con la noción de profundidad estadística, desarrollada inicialmente para el ordenamiento de datos multivariados. Considere en primer lugar la recta de números reales. Sean  $x_1$  y  $x_2$  dos números en  $\mathbb{R}$ , dado que este es un conjunto ordenado, se puede establecer la relación de orden entre  $x_1$  y  $x_2$  (bien sea  $>$ ,  $<$  ó  $=$ ). Por el contrario, sean  $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^p$  ( $p \geq 2$ ), no es posible establecer si  $\mathbf{x}_1 < \mathbf{x}_2$  o viceversa. Razón por la cual, se define la noción de profundidad para medir la centralidad de una observación con respecto al conjunto general de observaciones. Una de las medidas de profundidad más populares en el contexto multivariado, es la profundidad de Mahalanobis, definida por [Liu \(1990\)](#), basada en la distancia estadística que tiene el mismo nombre. Sea  $\mathbf{X}$  un vector aleatorio definido en  $\mathbb{R}^p$  con función de distribución  $F$ , la profundidad de Mahalanobis de una observación  $\mathbf{x}$  con respecto

a  $F$ , se define como:

$$MD(F; \mathbf{x}) = \frac{1}{1 + (\mathbf{x} - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})},$$

donde  $\boldsymbol{\mu} = E(\mathbf{X})$  y  $\boldsymbol{\Sigma}$  es la matriz de covarianzas asociadas al mismo proceso. En el caso de una muestra aleatoria  $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ ,  $\mathbf{x}_i \in \mathbb{R}^p$ , la profundidad de Mahalanobis muestral se define como  $MD(F_n, \mathbf{x}) = 1 / (1 + (\mathbf{x} - \bar{\mathbf{x}})^t \mathbf{S}^{-1} (\mathbf{x} - \bar{\mathbf{x}}))$ , siendo  $\bar{\mathbf{x}}$  y  $\mathbf{S}$  las estimaciones de  $\boldsymbol{\mu}$  y  $\boldsymbol{\Sigma}$ . Claramente, si una observación  $\mathbf{x}_i$  se aleja de la media muestral, tendrá una distancia estadística grande y en consecuencia, una profundidad pequeña. Por el contrario, si la observación está cerca del vector de medias  $\bar{\mathbf{x}}$ , la distancia estadística será pequeña y la medida de profundidad grande. En ese sentido, se dice que la observación con mayor profundidad es la observación más central del conjunto de datos.

Ahora, el concepto de *outlyingness* es inverso al de profundidad. [Dai and Genton \(2018\)](#), por ejemplo, definen una medida de *outlyingness* como  $o = 1/d - 1$ , siendo  $d(\cdot)$  una medida de profundidad. En ese sentido, la distancia de Mahalanobid puede verse como una medida de *outlyingness*, y puede interpretarse como un medida de la lejanía de una observación con respecto a un conjunto de datos. A continuación se presentan las propuestas de [Rousseeuw and Hubert \(2018\)](#) y [Dai and Genton \(2018\)](#), para la identificación de *outliers* en datos funcionales basadas en este concepto.

### 3.4.1. Directional Outlyingness

Siguiendo a [Rousseeuw and Hubert \(2018\)](#), para el caso univariado suponga una muestra ordenada  $Y = \{y_1, \dots, y_n\}$ . Es decir que  $y_i \leq y_j$  para todo  $i < j$ . Sea  $h = \lceil \lfloor \frac{h+1}{2} \rfloor \rceil$ , donde  $\lceil \cdot \rceil$  denota la parte entera, se definen  $Y_b = \{y_1, \dots, y_h\}$  y  $Y_a = \{y_h, \dots, y_n\}$ . Se computan los estimadores de escala

$$S_{0,a}(Y) = \frac{\text{med}(Z_a)}{\Phi^{-1}(0.75)} \quad \text{y} \quad S_{0,b}(Y) = \frac{\text{med}(Z_b)}{\Phi^{-1}(0.75)},$$

donde  $Z_a = Y_a - \text{med}(Y)$  y  $Z_b = Y_b - \text{med}(Y)$ . La medida DO para una observación  $y$  con respecto a la muestra  $Y$ , se define como

$$DO(y; Y) = \begin{cases} \frac{y - \text{med}(Y)}{S_a(Y)} & \text{si } y \geq \text{med}(Y) \\ \frac{\text{med}(Y) - y}{S_b(Y)} & \text{si } y \leq \text{med}(Y) \end{cases}, \quad (3-17)$$

donde

$$S_a(Y) = S_{0,a}(Y) \sqrt{\frac{1}{2\alpha h} \sum_{z_i \in Z_a} \rho_c \left( \frac{z_i}{S_{0,a}(Y)} \right)} \quad \text{y} \quad S_b(Y) = S_{0,b}(Y) \sqrt{\frac{1}{2\alpha h} \sum_{z_i \in Z_b} \rho_c \left( \frac{z_i}{S_{0,b}(Y)} \right)},$$

siento  $\alpha = \int_0^\infty \rho_c(x) d\Phi(x)$  y  $\rho_c(t) = \left(\frac{t}{c}\right)^2 \chi_{[-c,c]} + \chi_{(-\infty, c) \cup (c, \infty)}$  la función  $\rho$  de Huber para escala, de parámetro  $c$ .

Dada la muestra  $Y$ , para determinar si una observación es un *outlier* a partir de los valores  $\{DO_1, \dots, DO_n\}$ , se calculan los valores  $\{LDO_1, \dots, LDO_n\}$  con  $LDO_i = \log(0.1 + DO_i)$ . Esta transformación se realiza dado que la distribución de los valores  $DO$  presenta asimetría a derecha, por lo que se pretende obtener una distribución que se aproxime a la normal. Así, se dice que la observación  $y_i$  es un *outlier* si

$$\frac{LDO_i - \text{med}(LDO)}{\text{MAD}(LDO)} > \Phi^{-1}(0.995)$$

con  $\text{MAD}(Y) = \text{med}_i(|y_i - \text{med}_j(y_j)|) / \Phi^{-1}(0.75)$ , donde  $\Phi(\cdot)$  corresponde a la función de distribución acumulada de una variable aleatoria con distribución normal estándar. [Rousseeuw and Hubert \(2018\)](#) muestran que el  $DO$  funciona bien en la detección de observaciones atípicas incluso cuando la distribución de los datos presenta algún tipo de asimetría.

Ahora, para el caso funcional consideremos un conjunto de imágenes  $Y = \{Y_1, \dots, Y_n\}$  de  $J \times K$  pixeles. Se defien las medidas  $fDO$  y  $vDO$  para la  $i$ -ésima imagen en la muestra:

$$fDO(Y_i; Y) = \sum_{j=1}^J \sum_{k=1}^K DO(Y_i(j, k); Y(j, k)) \quad (3-18)$$

$$vDO(Y_i, Y) = \frac{\text{stdev}_{j,k}(DO(Y_i(j,k); Y(j,k)))}{1 + fDO(Y_i, Y)} \quad (3-19)$$

A diferencia del caso univariado o multivariado en el que los elementos atípicos se identifican como observaciones con valores extremos, en datos funcionales una observación puede ser atípica en magnitud, si dentro del dominio de la función presenta valores extremos, o puede ser un *outlier* de forma, si el comportamiento en el dominio difiere del comportamiento general de las funciones de la muestra, aún cuando los valores de la función no son extremos. En ese sentido, el  $fDO$  permite encontrar *outliers* de magnitud, mientras que el  $vDO$  permite identificar *outliers* de forma.

Para determinar el umbral a partir del cual una imagen se considera un *outlier*, se calcula la medida  $CFO$  (*combined fuctional outlyingness*):

$$CFO_i(Y_i, Y) = \sqrt{\left(\frac{fDO_i}{\text{med}(fDO)}\right)^2 + \left(\frac{vDO_i}{\text{med}(vDO)}\right)^2} \quad (3-20)$$

Análogo al caso univariado, una imagen  $Y_i$  es un *outlier* si

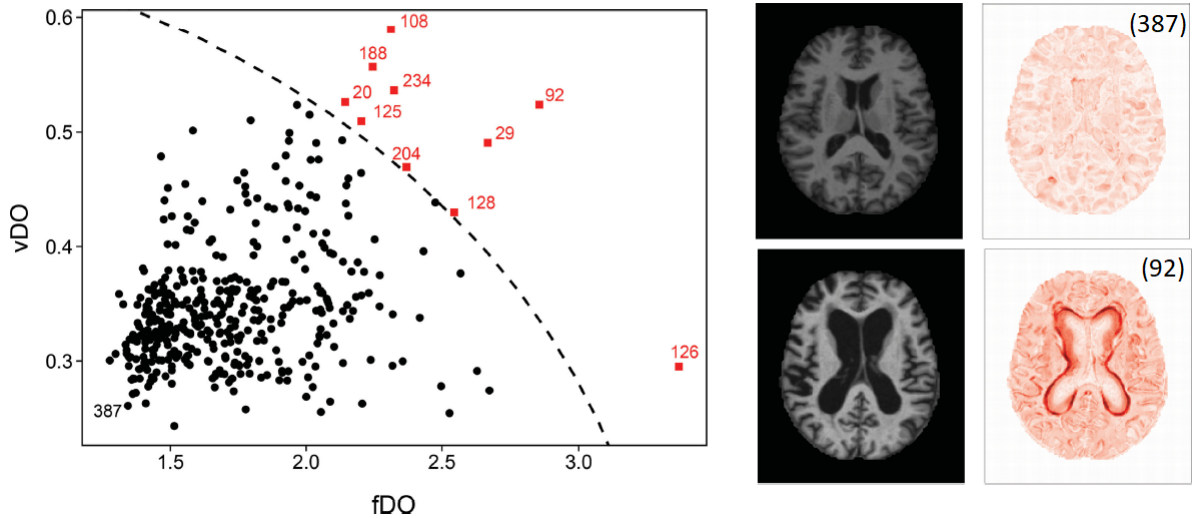
$$Z_{DO_i} = \frac{LCFO_i - \text{med}(LCFO)}{\text{MAD}(LCFO)} > \Phi^{-1}(0.995), \quad LCFO_i = \log(0.1 + CFO_i) \quad (3-21)$$

Ahora, teniendo en cuenta que para el cálculo de la medida  $fDO$  se calcula el  $DO$  punto por punto en todo el dominio de las imágenes, es posible construir mapas de calor con los valores  $DO$  para cada imagen estimada. Esto permite identificar las regiones específicas causantes de que una observación  $i$  sea identificada como *outlier*. A modo de ilustración, en la Figura (3-4), tomada de [Rousseeuw and Hubert \(2018\)](#), se muestra el *functional outlier*

$map$  (FOM) formado por los puntos  $fDO_i$  y  $vDO_i$  para una muestra compuesta por 413 imágenes de resonancia magnética (MRI). Sea  $k$  el valor de  $LCFO$  para el cual  $(LCFO - med(LCFO))/MAD(LCFO) = \Phi^{-1}(0.995)$  y sea  $b = e^k - 0.1$ , la línea punteada en el FOM presentado en la Figura (3-4) corresponde a la elipse conformada por los puntos  $(x, y)$  que satisfacen la siguiente ecuación:

$$b^2 = \left( \frac{x}{med(fDO)} \right)^2 + \left( \frac{y}{med(vDO)} \right)^2$$

Esta elipse permite ver gráficamente el umbral a partir del cual una observación se considera un *outlier*. Así, los puntos de color rojo marcan las observaciones identificadas como atípicas, dado que  $Z_{DO_i} > \Phi^{-1}(0.995)$ . En la Figura también se muestran las imágenes MRI originales en escala de grises y los mapas de calor de valores  $DO$  asociados para una imagen identificada como *outlier* (imagen 92) y la imagen más central de acuerdo al valor  $CFO$  (imagen 387). En el mapa de calor de la imagen 92 se identifican en color rojo oscuro las regiones causantes de que la observación sea identificada como *outlier*.



**Figura 3-4.:** FOM de muestra de imágenes MRI y mapas de calor de valores  $DO$  de las imágenes 387 y 92. Tomado de [Rousseeuw and Hubert \(2018\)](#).

Como se mencionó anteriormente, en este trabajo se propone utilizar este método para la identificación de observaciones influyentes en el contexto de los modelos GSIRM-TV, aplicando el método descrito sobre el conjunto de estimaciones  $\{\hat{\beta}_{(1)}, \dots, \hat{\beta}_{(n)}\}$ , siendo  $\hat{\beta}_{(i)}$  la estimación del parámetro imagen  $\beta$  obtenida omitiendo la  $i$ -ésima observación de la muestra. Así, se identifica a la observación  $i$ -ésima como una observación influyente si  $Z_{DO_i} > \Phi^{-1}(0.995)$ . Es decir, si la estimación  $\hat{\beta}_{(i)}$  es una estimación significativamente diferente del conjunto general de estimaciones  $\{\hat{\beta}_{(1)}, \dots, \hat{\beta}_{(n)}\}$ . Ahora, con el propósito de verificar que es apropiado comparar los valores  $Z_{DO_i}$  con el cuantil de una distribución normal estándar, se lleva a

cabo un estudio de simulación generando valores de la estadística  $Z_{DO}$  a partir de muestras generadas del modelo escalar-imagen sin contaminación, siguiendo el proceso que se describe a continuación.

Considere el siguiente modelo escalar-imagen:

$$Y_i = \langle X_i, \beta \rangle + \epsilon_i,$$

donde  $\beta$  es el parámetro imagen, el cual puede tomar la forma de alguna de las imágenes de la Figura (3-5),  $X_i$  la covariable imagen asociada y  $\epsilon_i \sim N(0, 1)$  para todo  $i = 1, \dots, n$ . En cada replicación del escenario simulado se llevan a cabo los siguientes pasos:

1. Se genera una muestra de tamaño  $n = 100$  del modelo escalar-imagen descrito y se estima el parámetro imagen  $\beta$ .
2. Se omite cada observación de la muestra y se estima el parámetro imagen, obteniendo el conjunto de estimaciones  $\{\hat{\beta}_{(1)}, \dots, \hat{\beta}_{(n)}\}$ .
3. Se aplica el método propuesto por [Rousseeuw and Hubert \(2018\)](#) sobre el conjunto de  $n$  estimaciones y se calculan los valores  $Z_{DO_i}$ ,  $i = 1, \dots, n$ .
4. Se aplican las pruebas de normalidad Shapiro-Wilk, Jarque-Bera y Kolmogorov-Smirnov sobre el conjunto de datos  $\{Z_{DO_1}, \dots, Z_{DO_n}\}$  y se determina el  $p$ -valor asociado a cada prueba.



**Figura 3-5.:** Parámetro imagen  $\beta$ .

La Tabla (3-1) muestra la proporción de veces que cada prueba rechaza la hipótesis de normalidad de los valores  $Z_{DO_i}$ , en 1000 repeticiones.

$\beta_0$	Shapiro-Wilk		Jarque-Bera		Kolmogorov-Smirnov	
	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.1$
T	0.999	1.000	0.792	0.991	0.397	0.601
Triángulo	0.999	1.000	0.780	0.986	0.388	0.597
Cuadrado	0.991	1.000	0.780	0.984	0.381	0.587
Ajedrez	0.999	1.000	0.818	0.986	0.427	0.628

**Tabla 3-1.:** Proporción de veces que se rechaza la hipótesis nula de normalidad en 1000 repeticiones, bajo niveles de significancia  $\alpha = 0.05$  y  $\alpha = 0.1$ , usando las pruebas de normalidad Shapiro-Wilk, Jarque-Bera y Kolmogorov-Smirnov.

Los resultados de las simulaciones evidencian que en general, la estadística  $Z_{DO}$  no sigue una distribución normal, por lo cual compararla con el cuantil de una variable aleatoria normal estándar no resultaría adecuado para el método de identificación de observaciones influyentes. Razón por la cual, se propone el método alternativo basado en la técnica de remuestreo *bootstrap* que se describe a continuación:

1. Estimar los parámetros del modelo escalar-imagen a partir de la muestra de tamaño  $n$ .
2. Obtener las estimaciones del parámetro imagen,  $\hat{\beta}_{(i)}$ ,  $i = 1, 2, \dots, n$ , omitiendo cada una de las observaciones de la muestra.
3. Con el conjunto de estimaciones  $\{\hat{\beta}_{(i)}\}_{i=1}^n$ , calcular los valores  $Z_{DO_i}$   $i = 1, 2, \dots, n$ .
4. Generar una muestra con reemplazamiento de tamaño  $n$  de los valores  $Z_{DO_i}$  y calcular el cuantil 0.995 de la muestra,  $q_1$ .
5. Repetir el paso (4)  $B - 1$  veces.
6. Comparar los valores  $Z_{DO_i}$  con el cuantil empírico estimado vía bootstrap,  $q = \frac{1}{B} \sum_{b=1}^B q_b$

Dado que el procedimiento anterior puede generar un valor umbral demasiado exigente, puesto que el cuantil 0.995 corresponde, por lo general, al máximo de los valores  $Z_{DO_i}$ , se propone usar la estimación del cuantil 0.95, como alternativa. Puesto que el uso de la estimación bootstrap del cuantil 0.995 puede generar que el método sea poco efectivo en la identificación de observaciones influyentes.

### 3.4.2. MS-PLOT

Otro método propuesto en los últimos años para la identificación de *outliers* en conjuntos de datos funcionales es el MS-PLOT, propuesto por [Dai and Genton \(2018\)](#). Este, al igual



que el método presentado anteriormente, se basa en el concepto de *outlyingness*.

Considerando el caso multivariado, supongamos un vector aleatorio  $\mathbf{Y}$  con función de distribución  $F_{\mathbf{Y}}$ . El *outlyingness* se define como

$$\mathbf{O}(\mathbf{Y}, F_{\mathbf{Y}}) = \{1/d(\mathbf{Y}, F_{\mathbf{Y}})\} \cdot \mathbf{v}, \quad d(\mathbf{Y}, F_{\mathbf{Y}}) > 0,$$

siendo  $\mathbf{v}$  un vector unitario que va de la mediana de  $F_d(\mathbf{Y}, F_{\mathbf{Y}})$  en dirección a  $\mathbf{Y}$  y  $d(\cdot)$ , una medida de profundidad. Ahora, Sea  $\mathbf{X}(t)$  un proceso estocástico funcional multivariado de dimensión  $p$ , con función de distribución  $F_{\mathbf{X}}$  y dominio  $I$ , Dai and Genton (2019) definen el *outlyingness* direccional funcional ( $FO$ ) de la siguiente forma:

$$FO(\mathbf{X}, F_{\mathbf{X}}) = \int_I \|\mathbf{O}(\mathbf{X}(t), F_{\mathbf{X}(t)})\|^2 \omega(t) dt$$

con  $\omega(t)$  una función de pesos. Los autores muestran que la medida anterior se puede expresar como  $FO = \|\mathbf{MO}\|^2 + VO$ , con  $\mathbf{MO}(\mathbf{X}, F_{\mathbf{X}}) = \int_I \mathbf{O}(\mathbf{X}(t), F_{\mathbf{X}(t)}) \omega(t) dt$  y  $VO(\mathbf{X}, F_{\mathbf{X}}) = \int_I \|\mathbf{O}(\mathbf{X}(t), F_{\mathbf{X}(t)}) - \mathbf{MO}(\mathbf{X}, F_{\mathbf{X}})\|^2 \omega(t) dt$ . Las componentes del vector  $\mathbf{MO} \in \mathbb{R}^p$  permiten identificar *outliers* de magnitud, mientras que los valores  $VO$  permiten la identificación de *outliers* de forma.

El método MS-PLOT a través del cual se identifican las observaciones atípicas, consiste en graficar los puntos  $(\mathbf{MO}^t, VO)^t$ , si  $p \leq 2$ . Es decir, si el proceso estocástico funcional multivariado es de dimensión 2, el MS-PLOT se puede representar en el plano 3-dimensional, graficando los puntos  $(MO_1, MO_2, VO)$ . Si  $p > 2$ , Dai and Genton (2018) proponen graficar los puntos  $(\|\mathbf{MO}\|, VO)^t$ . En cuyo caso, se pierde la información de la dirección en la cual se desvía la observación funcional multivariada. Es decir, la componente en la que una observación *outlier* difiere respecto del conjunto total de observaciones.

A modo de ilustración, siguiendo el ejemplo presentado en Dai and Genton (2018), en la Figura (3-6) se muestran dos ejemplos del MS-PLOT para conjuntos de datos funcionales. El primer conjunto de datos corresponde a funciones univariadas (ver Figura (3-6)(a)). En colores se marcan curvas correspondientes a observaciones *outlier* y en la Figura (3-6)(b) se muestra el MS-PLOT correspondiente. En el ejemplo, se evidencia que la curva de color amarillo corresponde a un *outlier* de magnitud y la curva de color azul claro a un *outlier* tanto de magnitud como de forma. Estas observaciones presentan valores  $MO$  altos, en valor absoluto. Por otra parte, las curvas de color azul oscuro, negro, naranja y verde, corresponden a *outliers* de forma únicamente. En consecuencia, presentan valores  $MO$  similares a los de las curvas grises (no atípicas o *inliers*), y valores  $VO$  bastante altos, como se observa en la Figura (3-6)(b). Lo cual muestra que las medidas  $MO$  y  $VO$  son efectivas para la identificación de distintos tipos de *outliers*.



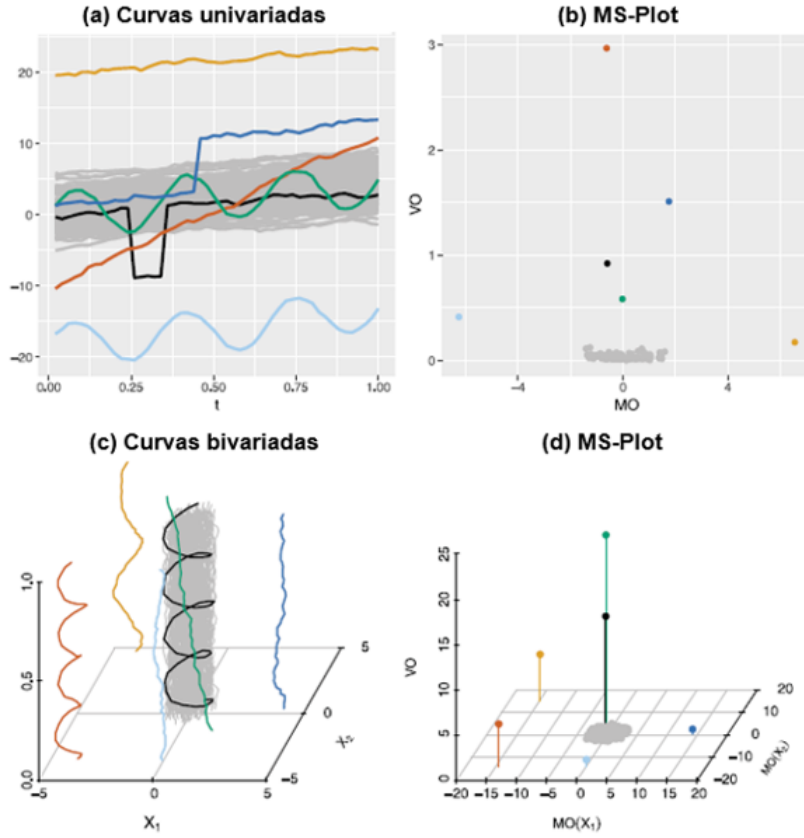


Figura 3-6.: MS-PLOT.

Por otra parte, el segundo conjunto de datos corresponde a curvas bivariadas (ver Figura (3-6)(c)). Análogo, al caso anterior, se marcan en colores las observaciones *outlier* y se construye el MS-PLOT asociado (Figura (3-6)(d)). En este ejemplo es evidente la importancia de graficar las componentes del vector  $\mathbf{MO}$ , puesto que permite identificar la dirección de la desviación en magnitud de las observaciones, como se señaló anteriormente. Por ejemplo, la curva de color azul oscuro, corresponde a un *outlier* de magnitud únicamente en la componente  $X_1$ , y por lo tanto, presenta un valor  $MO(X_1)$  alto y valores  $MO(X_2)$  y  $VO$  similares a los de las curvas grises (*inliers*). Adicionalmente, las curvas con forma de espiral, por ejemplo, corresponden a *outliers* de forma y presentan valores  $VO$  mayores a los de las curvas de color gris.

Adicional al MS-PLOT, Dai and Genton (2019) proponen un método para identificar las observaciones *outliers* que depende de un valor umbral contra el cual se compara una estadística calculada para cada observación del conjunto de datos. Considere un conjunto de  $n$  realizaciones,  $\{\mathbf{X}_1(s), \dots, \mathbf{X}_n(s)\}$ , de un proceso funcional multivariado  $\mathbf{X}(s)$  con dominio  $I$ , observadas en un conjunto finito de puntos  $S_k = \{s_1, \dots, s_k\}$  en  $I$ . Sean  $\mathbf{Y}_{k,n,i} = (\mathbf{MO}_{S_k,n,i}^t, VO_{S_k,n,i})$ ,  $i = 1, \dots, n$ , los valores  $MO$  y  $VO$  de las observaciones del conjunto de datos. Este método

consiste en calcular la siguiente estadística:

$$RMD_i^2 = (\mathbf{Y}_{k,n,i} - \bar{\mathbf{Y}}_{k,n,J}^*)^t \mathbf{S}_{k,n,J}^{-1} (\mathbf{Y}_{k,n,i} - \bar{\mathbf{Y}}_{k,n,J}^*) , \quad (3-22)$$

donde  $\bar{\mathbf{Y}}_{k,n,J}^* = h^{-1} \sum_{i \in J} \mathbf{Y}_{k,n,i}$ ,  $\mathbf{S}_{k,n,J}^* = h^{-1} \sum_{i \in J} (\mathbf{Y}_{k,n,i} - \bar{\mathbf{Y}}_{k,n,J}^*) (\mathbf{Y}_{k,n,i} - \bar{\mathbf{Y}}_{k,n,J}^*)^t$  y  $J$  el conjunto de  $h$  observaciones que minimizan el determinante de la matriz de covarianza asociada a los puntos  $\mathbf{Y}_{k,n,i}$ . Claramente, la estadística  $RMD_i^2$  corresponde al cuadrado de la distancia de Mahalanobis robusta a *outliers*, de los puntos  $(\mathbf{MO}^t, VO)^t$ , usando la estimación de la matriz de covarianzas por el método MCD (minimum covariance determinant), calculada por medio del algoritmo propuesto por [Rousseeuw and Driessen \(1999\)](#). Así, se dice que la  $i$ -ésima observación funcional es un *outlier* si se satisface la siguiente desigualdad

$$\frac{c(m-p)}{m(p+1)} RMD_i^2 > C$$

donde  $C$  es el cuantil  $\alpha$  de una distribución  $F_{p+1, m-p}$ , donde  $c$  y  $m$  son valores calculados vía simulación, como lo señalan [Dai and Genton \(2019\)](#).

## 4. Evaluación de la robustez a observaciones atípicas de modelos GSIRM-TV

En este capítulo se presentan los resultados de diferentes escenarios simulados para la evaluación la robustez de los modelos lineales generalizados escalar-imagen estimados vía variación total (GSIRM-TV), a la presencia de observaciones atípicas en la muestra. Adicionalmente, se evalúa el desempeño de los métodos propuestos para la identificación de observaciones influyentes. El capítulo está organizado en cuatro secciones: En la primera, se presentan los resultados de las simulaciones a través de las cuales se determina el nivel de afectación en la estimación de los parámetros, bajo diferentes tipos de contaminación en la muestra. En la segunda sección, se analiza la afectación del modelo en términos de su capacidad de predicción de nuevas observaciones. En la tercera, se determina la efectividad de los métodos propuestos para la identificación de observaciones influyentes, a través de un estudio de simulación. Finalmente, en la cuarta sección, se ilustra la aplicación de los métodos propuestos en un conjunto de datos reales.

### 4.1. Calidad de estimaciones de modelos GSIRM-TV ante presencia de observaciones atípicas

Como se mencionó anteriormente, la evaluación de la robustez de los modelos GSIRM-TV ante distintos tipos de contaminación en la muestra se centra en dos aspectos: El primero es el desempeño del método de estimación en términos de calidad de las estimaciones y el segundo es el desempeño del modelo en términos de predicción de nuevas observaciones. Para el estudio del primer aspecto, se utiliza la distancia inducida por la norma  $L_2$  para determinar la calidad de las estimaciones del vector de parámetros escalares  $\theta$ :

$$d(\theta, \hat{\theta}) = \|\theta - \hat{\theta}\|_2 = \left\{ \sum_{j=1}^p (\theta_j - \hat{\theta}_j)^2 \right\}^{1/2}. \quad (4-1)$$

Dado que en los escenarios simulados solo se tendrá en cuenta el intercepto como parámetro escalar, adicionalmente, se calcula el promedio del error cuadrático de las estimaciones del total de replicaciones de cada escenario,  $d_2(\theta, \hat{\theta}) = 1/R \sum_{k=1}^R (\theta - \hat{\theta}_k)^2$ , siendo  $R$  el número

de replicaciones y  $\hat{\theta}_k$  la estimación de  $\theta$  en la replicación  $k$ .

Para las estimaciones del parámetro imagen  $\beta$ , se hará uso de la distancia inducida por la norma TV descrita en la expresión (3-9):

$$d(\beta, \hat{\beta}) = \|\beta - \hat{\beta}\|_{TV} = \sum_{j,k} \left\{ |\nabla(\beta - \hat{\beta})_{jk,1}| + |\nabla(\beta - \hat{\beta})_{jk,2}| \right\}. \quad (4-2)$$

Adicionalmente, sea  $\beta_v = Vec(\beta) \in \mathbb{R}^{N^2}$  el parámetro imagen verdadero vectorizado y  $\hat{\beta}_v$  su estimación vectorizada, se hace uso de la distancia  $L_2$  para determinar la diferencia entre el valor verdadero del parámetro imagen y el valor estimado pixel a pixel:

$$d_2(\beta_v, \hat{\beta}_v) = \|\beta_v - \hat{\beta}_v\|_2 = \left\{ \sum_{j=1}^{N^2} (\beta_{v(j)} - \hat{\beta}_{v(j)})^2 \right\}^{1/2}, \quad (4-3)$$

donde  $\beta_{v(j)}$  es la componente  $j$ -ésima de  $\beta_v$  y  $N$  el número de pixeles de la imagen  $\beta$  en dirección horizontal y vertical. Cabe anotar que aunque en teoría,  $\beta$  pertenece al espacio de variación acotada, para efectos del estudio de simulación, esta es generada como una imagen compuesta por un número finito de pixeles ( $N \times N$ ). Razón por la cual es posible vectorizarla.

Para el estudio de simulación, se consideran los siguientes modelos de respuesta continua y binaria:

- Modelo de respuesta continua

$$Y_i = \theta_0 + \langle X_i, \beta_0 \rangle + \epsilon_i. \quad (4-4)$$

- Modelo de respuesta binaria

$$Y_i | (X_i, Z_i) \sim \text{Bernoulli}(\mu_i) \\ \ln \left( \frac{\mu_i}{1 - \mu_i} \right) = \theta_0 + \langle X_i, \beta_0 \rangle, \quad (4-5)$$

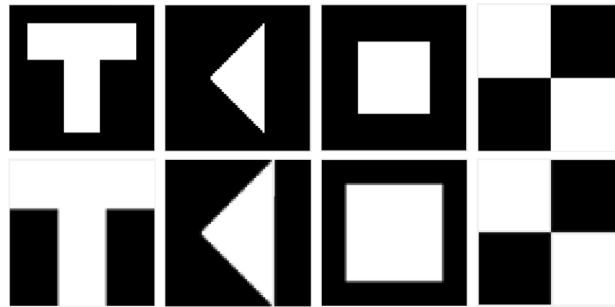
donde  $\theta_0 = 2$ ,  $\beta_0$  es una imagen dentro del conjunto presentado en la Figura (4-1). En el caso del modelo con respuesta continua,  $\epsilon_i \sim N(0, 1)$ .



Figura 4-1.: Parámetro imagen  $\beta$ .

Para el modelo contaminado se consideran los siguientes tipos de contaminación:

- **Cambio de magnitud:** Se generan observaciones de un modelo cuyo parámetro imagen es  $\beta_1 = \gamma_b \beta_0$ , con  $\gamma_b = \{1/5, 1/2, 2, 5\}$ . Esto quiere decir que las observaciones contaminadas provienen de un modelo en el que el parámetro imagen tiene la misma forma del parámetro imagen del modelo sin contaminación, pero cuyos valores de los pixeles difieren de los de la imagen original.
- **Cambio de forma - Tipo 1:** Se generan observaciones de un modelo cuyo parámetro imagen es  $\beta_1 \neq \gamma_b \beta_0$ . Es decir,  $\beta_1$  es una imagen del conjunto presentado en la Figura (4-1) diferente a la imagen  $\beta_0$ .
- **Cambio de forma - Tipo 2:** Para el modelo contaminado, se generan observaciones en los que el parámetro imagen ( $\beta_1$ ) es la misma imagen original ( $\beta_0$ ), pero sobre la cual se hace un aumento del 50 %. En la Figura (4-2) se muestran las imágenes del modelo original y las del modelo contaminado. Este tipo de cambio se estudia únicamente bajo el modelo de respuesta continua.



**Figura 4-2.:** Primera fila. Parámetro imagen  $\beta_0$ . Segunda fila: Parámetro imagen  $\beta_1$  (Modelo de contaminación).

Siguiendo a Wang et al. (2017), para el proceso de generación de los datos de acuerdo a los modelos de respuesta continua (4-4) y binaria (4-5), se generan las covariables imágenes haciendo uso de la base bivariada de Haar para  $L_2(\Omega)$ . Considere las siguientes funciones univariadas:

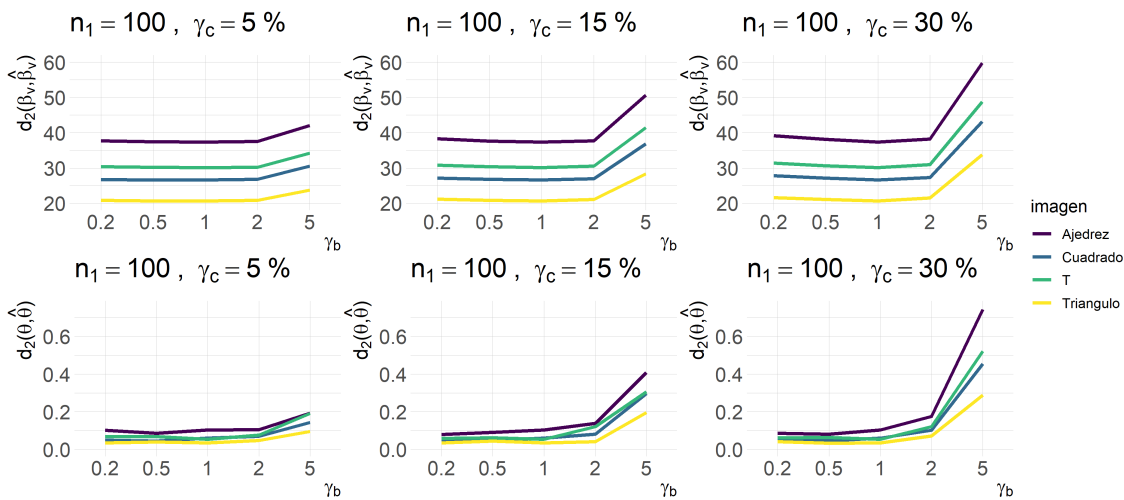
$$\begin{aligned} \phi^0(t) &= I_{[0,1)} \\ \phi^1(t) &= \begin{cases} 1 & \text{si } t \in [0, 1/2) \\ -1 & \text{si } t \in [1/2, 1) \end{cases} \end{aligned}$$

Se construyen funciones con dominio en  $\mathbb{R}^2$ ,  $\phi^d(s, t) = \phi(s)^{d_1} \phi(t)^{d_2}$ , con  $d = \{(0, 1), (1, 0), (1, 1)\}$ . La base bivariada de Haar restringida a  $N^2$  funciones base está conformada por el conjunto  $\{I_{[0,1)^2}, \phi_{j,k}^d\}$  con  $j \leq J - 1$ , siendo  $J = \log_2(N)$  y  $N$  la dimensión de las imágenes. Las

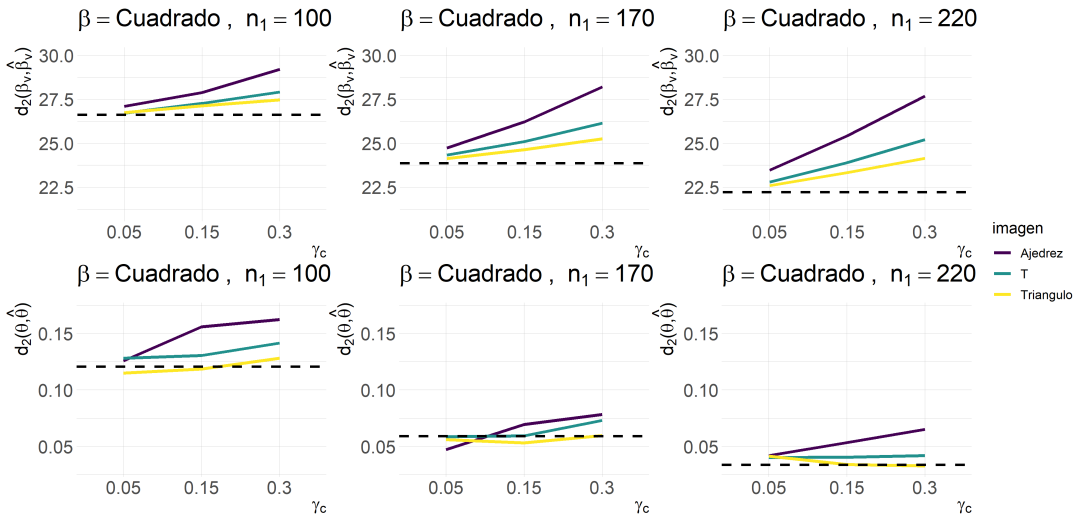
covariables son generadas del proceso  $X_i = \sum_l \xi_{il} \phi_l$ , donde  $\phi_l$  es la componente  $l$ -ésima de la base bivariada descrita y  $\xi_{il}$  una variable aleatoria con distribución normal estándar.

En cada replicación de los escenarios simulados, se generan  $n_1$  observaciones para el conjunto de entrenamiento, de las cuales  $n_c = \lceil \gamma_c n_1 \rceil$  corresponden a observaciones que provienen del modelo contaminado, donde  $\lceil \cdot \rceil$  es la función parte entera y  $\gamma_c = \{5\%, 15\%, 30\%\}$  corresponde al nivel de contaminación en la muestra de entrenamiento. Para el conjunto de prueba se generan  $n_2 = \lceil 0.3 n_1 \rceil$  observaciones del modelo sin contaminación. Se consideran tamaños de la muestra de entrenamiento,  $n_1 = 100, 170, 220$ . De cada escenario se realizan  $R = 200$  replicaciones y se calculan el promedio y desviación estándar de los valores  $d(\beta, \hat{\beta})$ ,  $d_2(\beta_v, \hat{\beta}_v)$ ,  $d(\theta, \hat{\theta})$ . Adicionalmente, se calcula el valor  $d_2(\theta, \hat{\theta})$  para el total de replicaciones, como se mencionó anteriormente. A continuación se presentan los resultados de los escenarios simulados.

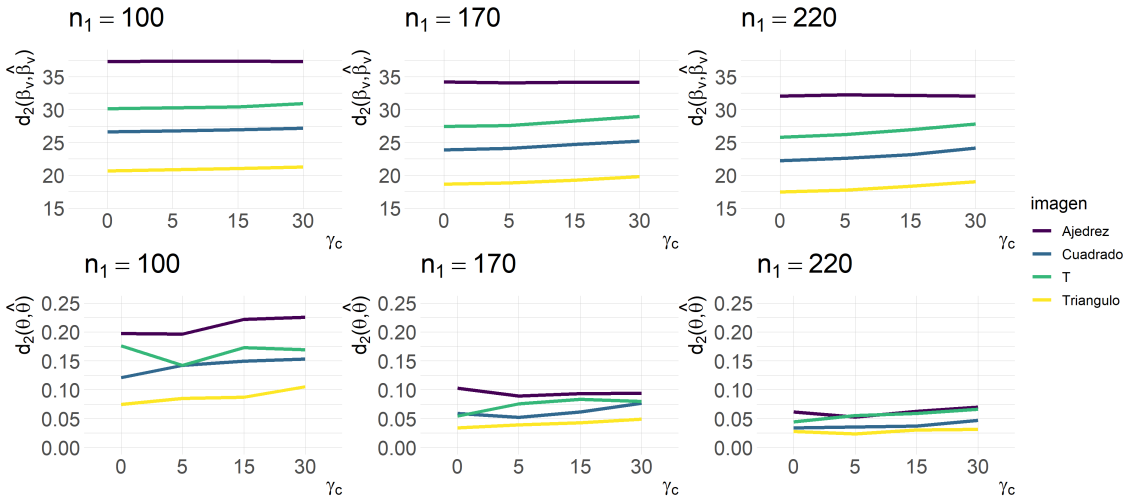
En la Figura (4-3) se observan los valores promedio de la medida  $d_2(\beta_v, \hat{\beta}_v)$  y los valores de  $d_2(\theta, \hat{\theta})$ , bajo el modelo de respuesta continua, para los distintos factores de contaminación de magnitud  $\gamma_b$  y para cada tipo de imagen  $\beta_0$  (Ajedrez = morado, Cuadrado = azul, Triángulo = amarillo y T = verde), bajo un tamaño de muestra de entrenamiento de  $n_1 = 100$ . La Figura sugiere que la contaminación de magnitud afecta negativamente tanto la estimación del parámetro imagen como la estimación del intercepto del modelo, en especial cuando el factor de contaminación es mayor a 1. Los resultados con la medidas  $d(\beta, \hat{\beta})$  y  $d(\theta, \hat{\theta})$  y bajo otros tamaños de muestra son similares, como se puede observar en las Tablas (A-1), (A-2) y (A-3) del Anexo A.



**Figura 4-3.:** Valores promedio  $d_2(\beta_v, \hat{\beta}_v)$  y valores  $d_2(\theta, \hat{\theta})$ . Modelo de respuesta continua, bajo niveles de contaminación  $\gamma_c$  (%) y factor de contaminación de magnitud  $\gamma_b$ .  $\theta_0 = 2$ ,  $n = 100$ .



**Figura 4-4.:** Valores promedio  $d_2(\beta_v, \hat{\beta}_v)$  y valores  $d_2(\theta, \hat{\theta})$ . Modelo de respuesta continua, bajo niveles de contaminación  $\gamma_c$  (%) y contaminación de cambio de forma - Tipo 1.  $\theta_0 = 2$ ,  $\beta_0 = \text{Cuadrado}$ . Línea punteada indica valores promedio de  $d_2(\beta_v, \hat{\beta}_v)$  y valores de  $d_2(\theta, \hat{\theta})$ , bajo el escenario sin contaminación.

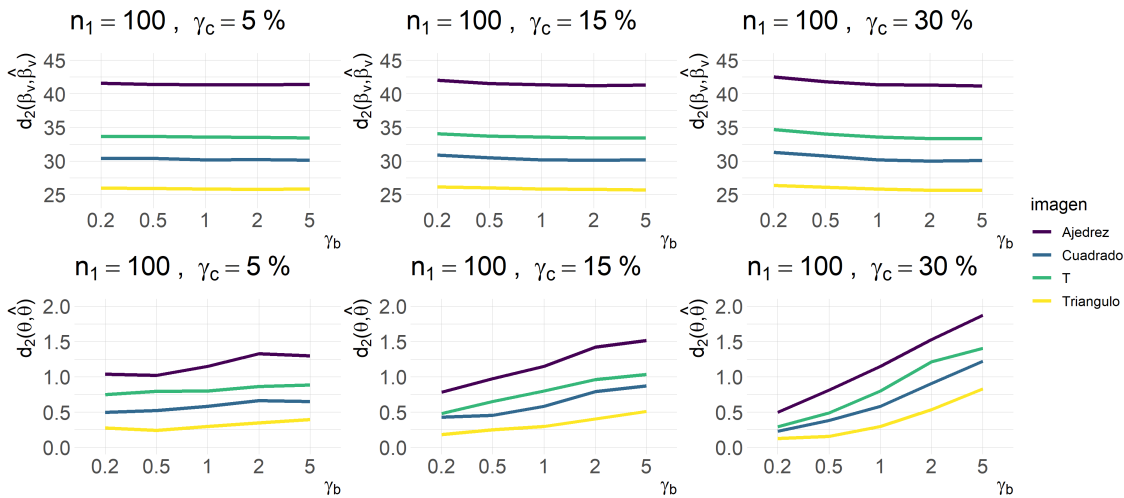


**Figura 4-5.:** Valores promedio  $d_2(\beta_v, \hat{\beta}_v)$  y valores  $d_2(\theta, \hat{\theta})$ . Modelo de respuesta continua, bajo niveles de contaminación  $\gamma_c$  (%) y contaminación de cambio de forma - Tipo 2.

Por otra parte, en la Figura (4-4) se observan los valores promedio de la medida  $d_2(\beta_v, \hat{\beta}_v)$  y los valores de  $d_2(\theta, \hat{\theta})$ , bajo el modelo de respuesta continua y contaminación de cambio de forma - Tipo 1 en el parámetro imagen, donde la imagen original ( $\beta_0$ ) tiene forma de Cuadrado. La Figura muestra que bajo este tipo de contaminación se afecta de forma negativa tanto la estimación del parámetro imagen, como la del intercepto. Este detrimento de las

estimaciones se hace más evidente cuando aumenta el nivel de contaminación en la muestra. Adicionalmente, se observa una clara relación entre el nivel de afectación y el área de la imagen del modelo de contaminación, puesto que los valores más altos se presentan cuando la imagen del modelo de contaminación corresponde a un Ajedrez (imagen de mayor área), seguida de las imágenes con forma de T y Triángulo. Cabe anotar, que bajo este escenario, ninguno de los valores  $d_2(\beta_v, \hat{\beta}_v)$  y  $d_2(\theta, \hat{\theta})$ , alcanza los niveles observados en la Figura (4-3). Lo cual indica que la contaminación de cambio de forma - Tipo 1 afecta las estimaciones en menor medida que los cambios de magnitud. Los resultados bajo los tipos de imágenes restantes son similares, como se puede ver en las Tablas (A-7), (A-8) y (A-9) del Anexo A.

Por último, en la Figura (4-5) se muestran los valores promedio de la medida  $d_2(\beta_v, \hat{\beta}_v)$  y los valores de  $d_2(\theta, \hat{\theta})$ , bajo el modelo de respuesta continua y contaminación de cambio de forma - Tipo 2. En general se observa que este tipo de contaminación no afecta de forma importante las estimaciones del parámetro imagen y del intercepto, aún cuando el nivel de contaminación en la muestra es grande ( $\gamma_c = 30\%$ ).



**Figura 4-6.:** Valores promedio  $d_2(\beta_v, \hat{\beta}_v)$  y valores  $d_2(\theta, \hat{\theta})$ . Modelo de respuesta binaria, bajo niveles de contaminación  $\gamma_c$  (%) y factor de contaminación de magnitud  $\gamma_b$ .  $\theta_0 = 2$ ,  $n = 100$ .

Ahora, en la Figura (4-6) se presentan los valores promedio de la medida  $d_2(\beta_v, \hat{\beta}_v)$  y los valores de  $d_2(\theta, \hat{\theta})$  para el caso del modelo de respuesta continua, bajo contaminación de magnitud y un tamaño de la muestra de entrenamiento de  $n_1 = 100$ . La Figura muestra que este tipo de contaminación afecta principalmente la estimación del intercepto. La estimación del parámetro imagen no parece verse afectada de forma importante. En la Figura se observa que para valores  $\gamma_b$  menores a 1, la medida  $d_2(\beta_v, \hat{\beta}_v)$  aumenta levemente, indicando un deterioro de la calidad de estimación del parámetro imagen, mientras que para valores de  $\gamma_b$



mayores a 1, la medida  $d_2(\beta_v, \hat{\beta}_v)$  presenta valores muy similares al escenario sin contaminación. Contrario a lo que sucede con el modelo de respuesta continua. Por otro lado, bajo contaminación de cambio de forma - Tipo 1, se observa un deterioro de las estimaciones del parámetro imagen y del intercepto, como se observa en la Figura (4-7), bajo el escenario con imagen parámetro  $\beta_0$  con forma de Cuadrado. Al igual que en el caso del modelo de respuesta continua, la afectación es mayor a medida que aumenta el nivel de contaminación  $\gamma_c$  y nuevamente se observa la relación entre el área de la imagen del modelo de contaminación y los valores de las medidas de calidad de las estimaciones, siendo mayor el error de estimación cuanto mayor es el área de la imagen del modelo de contaminación. Los resultados bajo los demás tipos de imágenes bajo el modelo respuesta binaria se pueden consultar en las tablas Tablas (A-10), (A-11), y (A-12) del Anexo A.

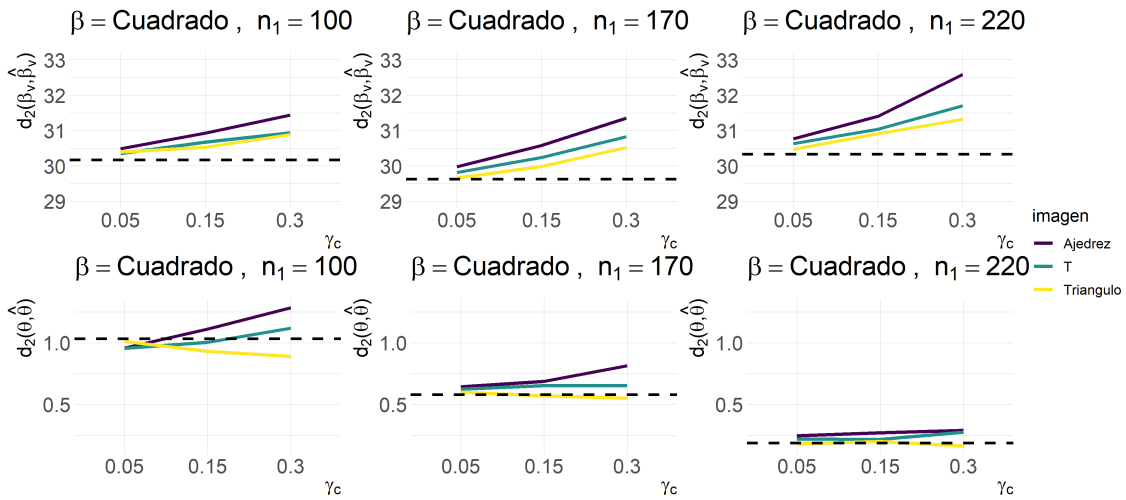


Figura 4-7.: Valores promedio  $d_2(\beta_v, \hat{\beta}_v)$  y valores  $d_2(\theta, \hat{\theta})$ . Modelo de respuesta binaria, bajo niveles de contaminación  $\gamma_c$  (%) y contaminación de cambio de forma - Tipo 1.  $\theta_0 = 2$ ,  $\beta_0 = \text{Cuadrado}$ .

## 4.2. Predicción de nuevas observaciones en modelos GSIRM-TV ante presencia de observaciones atípicas

En cuanto al segundo aspecto a considerar en la evaluación de la robustez del modelo bajo escenarios de de contaminación en la muestra, a continuación se especifican las métricas de nivel de predicción para modelos de respuesta continua y binaria:

1. Modelo de respuesta continua

Considere de forma general el siguiente modelo escalar-función:

$$Y_i = \theta^t Z_i + \langle X_i, \beta \rangle + \epsilon_i, \quad (4-6)$$

donde  $\beta$  es el parámetro imagen,  $X_i$  la covariable imagen asociada,  $Z_i \in \mathbb{R}^p$  las covariables escalares y  $\theta$  el vector de parámetros asociados.

Sea  $\{(Y_{1,1}, X_{1,1}, Z_{1,1}), \dots, (Y_{n_1,1}, X_{n_1,1}, Z_{n_1,1})\}$  el conjunto de entrenamiento del modelo de tamaño  $n_1$ . Es decir, el conjunto de datos con el cual se estiman los parámetros  $\beta$  y  $\theta$ . Sea  $\{(Y_{1,2}, X_{1,2}, Z_{1,2}), \dots, (Y_{n_2,2}, X_{n_2,2}, Z_{n_2,2})\}$  el conjunto de prueba de tamaño  $n_2$ . Para evaluar la calidad de predicción del modelo, se calcula la raíz del error cuadrático medio de predicción (RMSPE) de la siguiente forma:

$$RMSPE = \sqrt{\frac{1}{n_2} \sum_{i=1}^{n_2} (Y_{i,2} - \hat{Y}_{i,2})^2}, \quad (4-7)$$

donde  $\hat{Y}_{i,2} = \hat{\theta}^t Z_{i,2} + \langle X_{i,2}, \hat{\beta} \rangle$ . Lógicamente, si el modelo es adecuado y las estimaciones de los parámetros son buenas, se espera que el modelo prediga bien la variable respuesta de las observaciones en el conjunto de prueba, y en consecuencia se espera un valor  $RMSPE$  pequeño. Caso contrario si las estimaciones de  $\beta$  y  $\theta$  no son buenas estimaciones.

## 2. Modelo de respuesta binaria

Considere el siguiente modelo lineal generalizado escalar-imagen de respuesta binaria:

$$Y_i | (X_i, Z_i) \sim \text{Bernoulli}(\mu_i)$$

$$\ln \left( \frac{\mu_i}{1 - \mu_i} \right) = \theta^t Z_i + \langle X_i, \beta \rangle,$$

Sea 1 el valor de la variable respuesta para los elementos de la clase positiva y 0 el valor de la variable respuesta para los elementos de la clase negativa, para evaluar el desempeño del modelo de clasificación binaria, se hace uso de la curva ROC (*receiver operating characteristic*). La curva ROC asociada al modelo se construye a partir de los conteos de elementos del conjunto de prueba que se clasificaron correctamente y los que no, dados el modelo estimado y el valor umbral definido  $\alpha$  para el cual se clasifica a un elemento en la clase positiva si  $\hat{\mu} > \alpha$ , siendo  $\mu = P(\{Y = 1\})$ , la probabilidad de que dicho elemento pertenezca a la clase positiva. Teniendo en cuenta lo anterior, se definen los siguientes conceptos necesarios para la construcción de la curva ROC:

- Verdadero Positivo (VP): Elementos de la clase positiva clasificados correctamente.
- Verdadero Negativo (VN): Elementos de la clase negativa clasificados correctamente.

- Falso Positivo (FP): Elementos de la clase negativa clasificados incorrectamente.
- Falso Negativo (FN): Elementos de la clase positiva clasificados incorrectamente.

Sea  $n_2$  el tamaño del conjunto de prueba del modelo de clasificación binaria, los resultados de la clasificación en dicho conjunto se resume en la matriz de confusión descrita a continuación:

		Valor predicho	
		1	0
Valor verdadero	1	VP	FN
	0	FP	VN

**Tabla 4-1.:** Matriz de confusión para un modelo de clasificación binaria

Claramente  $n_2 = VP + FP + FN + VN$ . Ahora, para evaluar si el modelo clasifica bien en ambas clases se calculan las siguientes medidas:

- Tasa de verdaderos positivos (TVP) o sensibilidad: Se define como la razón entre los verdaderos positivos y el total de elementos de la clase positiva en el conjunto de prueba.

$$TVP = \frac{VP}{VP + FN} \tag{4-8}$$

- Tasa de verdaderos negativos (TVN) o especificidad: Se define como la razón entre los verdaderos negativos y el total de elementos de la clase negativa en el conjunto de prueba.

$$TVN = \frac{VN}{VN + FP} \tag{4-9}$$

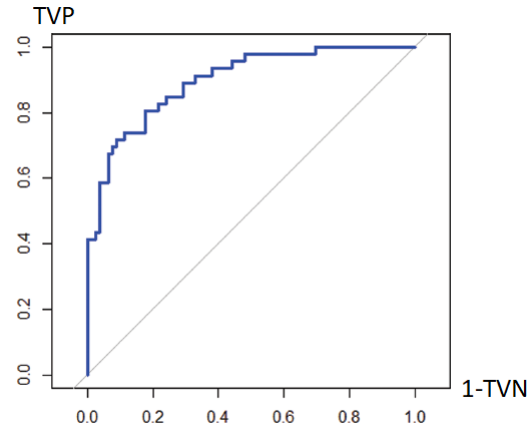
- Tasa de falsos negativos (TFN): Se define como la razón entre los falsos negativos y el total de elementos de la clase positiva en el conjunto de prueba.

$$TFN = \frac{FN}{VP + FN} \tag{4-10}$$

- Tasa de falsos positivos (TFP): Se define como la razón entre los falsos positivos y el total de elementos de la clase negativa en el conjunto de prueba.

$$TFP = \frac{FP}{VN + FP} \tag{4-11}$$

La curva ROC se obtiene graficando los valores  $1 - TVN$  contra la sensibilidad ( $TVP$ ), para varios valores umbral  $\alpha$ . En la Figura (4-8) se presenta un ejemplo de una curva ROC:



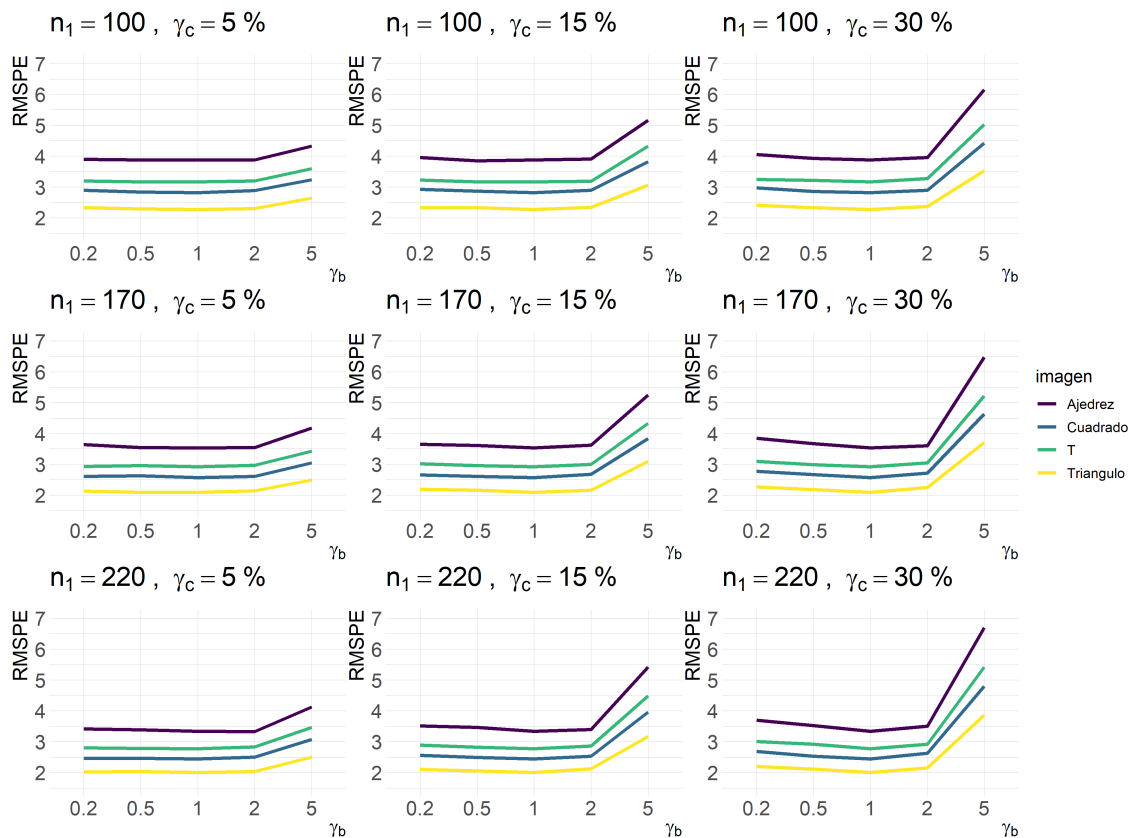
**Figura 4-8.:** Curva ROC asociada a modelo de clasificación binaria. Tomado de [Yang and Berdine \(2017\)](#)

Ahora, si el modelo de clasificación es perfecto para todo  $\alpha$ , es decir, clasifica correctamente todos elementos del conjunto de prueba, entonces la sensibilidad y la especificidad serán iguales a 1 y la curva ROC será el punto  $(0, 1)$ . Si la sensibilidad es igual a la tasa de falsos negativos ( $TFN$ ), se dice que el clasificador no tiene utilidad, puesto que se desempeña igual que un clasificador aleatorio. En ese caso la curva ROC tomará valores en el segmento del punto  $(0, 0)$  al punto  $(1, 1)$ . Usualmente las curvas ROC asociadas a modelos de clasificación binaria son curvas convexas ( $TVP > 1 - TVN$ ). Entre más convexa sea la curva, mejor será el clasificador. Una medida que resume el desempeño del clasificador es el área bajo la curva ROC ( $ABC$ ). Si la curva ROC es la diagonal del punto  $(0, 0)$  al punto  $(1, 1)$ , el valor  $ABC$  será 0.5. Luego, el valor  $ABC$  se aproximará a 1 si el clasificador tiene buen desempeño.

Los resultados de las simulaciones se obtienen de los escenarios descritos en la sección anterior. Es decir que en cada replicación, se generan  $n_1$  observaciones para el conjunto de entrenamiento, incluyendo  $n_c = \lceil \gamma_c n_1 \rceil$  observaciones del modelo contaminado, con niveles de contaminación  $\gamma_c = \{5\%, 15\%, 30\%\}$ . Para el conjunto de prueba se generan  $n_2 = \lceil 0.3n_1 \rceil$  observaciones del modelo sin contaminación. Se consideran tamaños de la muestra de entrenamiento,  $n_1 = 100, 170, 220$ . De cada escenario se realizan  $R = 200$  replicaciones y se calculan el promedio y desviación estándar de los valores,  $RMSPE$  para el modelo de respuesta continua y  $ABC$  para el de respuesta binaria. A continuación se presentan los resultados de los escenarios simulados.

En la Figura (4-9) se muestran los valores  $RMSPE$  promedio de las 200 replicaciones, para los distintos factores de contaminación de magnitud  $\gamma_b$  y los tamaños de muestra considerados, para cada tipo de imagen  $\beta_0$  (Ajedrez = morado, Cuadrado = azul, Triángulo = amarillo

y T = verde). Se observa que en el caso del modelo de respuesta continua, los factores de contaminación  $\gamma_b = 1/5, 1/2$ , no generan una afectación grande en términos de predicción. Sin embargo, los factores de contaminación mayores sí afectan considerablemente el desempeño del modelo estimado, en especial para  $\gamma_b = 5$ . Adicionalmente, la afectación es mayor cuanto mayor es el porcentaje de contaminación en la muestra ( $\gamma_c$ ), lo cual se relaciona directamente con el hecho de que las estimaciones del parámetro imagen y del intercepto del modelo se afectan considerablemente bajo estas condiciones, como se observó en la sección anterior.



**Figura 4-9.:** Valores  $RMSPE$  promedio. Modelo de respuesta continua, bajo niveles de contaminación  $\gamma_c$  (%) y factor de contaminación de magnitud  $\gamma_b$ .  $\theta_0 = 2$ . Primera fila ( $n = 100$ ), segunda fila ( $n = 170$ ), tercera fila ( $n = 220$ ).

Por otra parte, en la Figura (4-10) se observa cómo se comportan los valores  $RMPSE$  promedio cuando aumenta el tamaño de muestra, cuando el factor de contaminación es  $\gamma_b = 2$ . Aquí se aprecia que la afectación en términos de predicción es menor a medida que aumenta el tamaño de muestra, independientemente del nivel de contaminación  $\gamma_c$ . Sin embargo, cuando  $\gamma_b = 5$  la afectación del modelo es mayor cuando aumenta el nivel de contaminación y el tamaño de muestra.

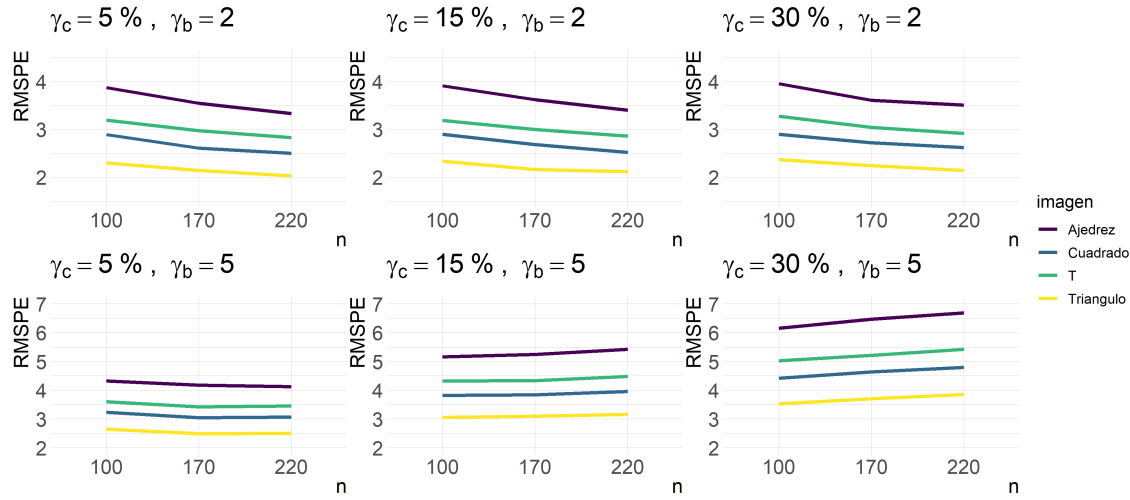


Figura 4-10.: Valores  $RMSPE$  promedio. Modelo de respuesta continua, bajo niveles de contaminación  $\gamma_c$  (%) y factor de contaminación de magnitud  $\gamma_b$ .  $\theta_0 = 2$ . Primera fila ( $\gamma_b = 2$ ), segunda fila ( $\gamma_b = 5$ ).

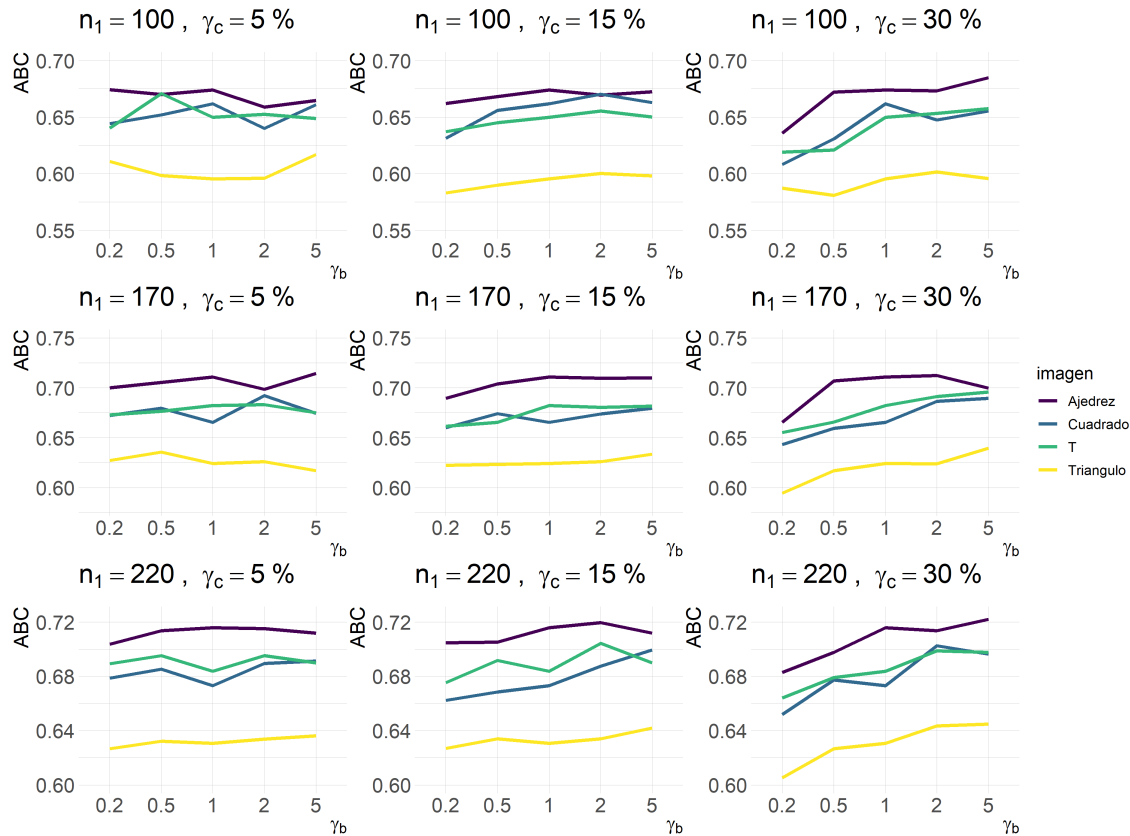


Figura 4-11.: Valores  $ABC$  promedio. Modelo de respuesta binaria, bajo niveles de contaminación  $\gamma_c$  (%) y factor de contaminación de magnitud  $\gamma_b$ .  $\theta_0 = 2$ . Primera fila ( $n = 100$ ), segunda fila ( $n = 170$ ), tercera fila ( $n = 220$ ).

Para el caso del modelo de respuesta binaria con función de enlace *logit*, no se observa una relación clara entre el factor de contaminación de magnitud ( $\gamma_b$ ) y la afectación en términos de la medida *ABC*, como se muestra en la Figura (4-11). En los escenarios considerados, los valores *ABC* fluctúan en niveles cercanos al escenario sin contaminación ( $\gamma_b = 1$ ). Lo cual es coherente con lo observado en la sección anterior, donde no se evidenció una afectación notoria en la estimación del parámetro imagen, pese a que la calidad de estimación del intercepto se ve altamente afectada. Por otro lado, en el caso del modelo de respuesta binaria, la predicción del modelo mejora cuando aumenta el tamaño de muestra, bajo todos los escenarios, como se evidencia en la Figura (4-12). Esto se debe a que mejoran las estimaciones del parámetro imagen y el parámetro escalar cuando aumenta el tamaño de muestra, como se observa en las Tablas (A-4), (A-5) y (A-6) del Anexo A, para valores  $\gamma_c$  y  $\gamma_b$  fijos.

Ahora, en cuanto a la afectación ante contaminación con observaciones que provienen de un modelo cuyo parámetro imagen tiene una forma distinta a la del modelo original (cambio de forma Tipo 1), en la Figura (4-13) se muestran los valores *RMSPE* promedio para los tamaños de muestra considerados y cada nivel de contaminación en la muestra. En primer lugar, se observa que en el caso del modelo de respuesta continua, cuanto mayor es el nivel de contaminación, mayor la afectación en términos de predicción. Adicionalmente, se evidencia que la mayor afectación se genera cuando la imagen original ( $\beta_0$ ) es una T, un Triángulo o un Cuadrado y la imagen del modelo contaminado ( $\beta_1$ ) es un Ajedrez. Esto se debe a que el área de la imagen de Ajedrez es la más grande. Por lo cual, afecta la estimación del parámetro imagen en un área mayor. Por otro lado, cuando la imagen  $\beta_0$  es un Ajedrez, no existe una diferencia evidente del nivel de afectación entre las demás imágenes, aunque sí existe un deterioro del modelo en términos de predicción, que se hace más evidente al aumentar el nivel de contaminación en la muestra.

Para el modelo de respuesta binaria, en general, la capacidad de predicción del modelo se ve afectada cuando el nivel de contaminación en la muestra aumenta, aunque la afectación es leve, especialmente en los escenarios en los que el área de la imagen original tiene forma de Triángulo, Cuadrado o T. Cuando el parámetro imagen del modelo original tiene forma de Ajedrez, la afectación de la predicción ante este tipo de cambios es más evidente, dado que el modelo es contaminado con observaciones de modelos con imagen parámetro de menor área, como puede verse en la Figura (4-14).

Finalmente, adicional a los tipos de contaminación estudiados hasta este punto, se consideran cambios de forma en el parámetro imagen, en los que la imagen del modelo contaminado es la misma imagen original, pero sobre la cual se hace un aumento del 50% (cambio de forma Tipo 2). En esta parte del estudio de simulación, solo se considera el modelo de respuesta continua, dado que los resultados anteriores muestran que cualquier tipo de cambio tiene poco efecto en la estimación de los parámetros del modelo de respuesta binaria y en consecuencia, baja afectación en la capacidad de predicción del mismo.

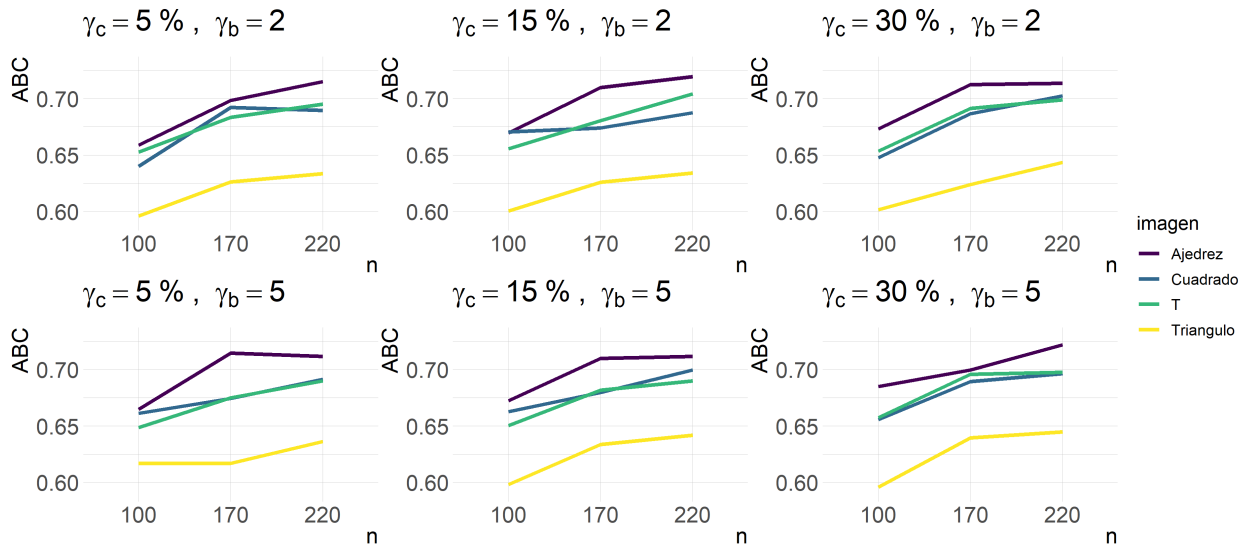


Figura 4-12.: Valores  $ABC$  promedio. Modelo de respuesta binaria, bajo niveles de contaminación  $\gamma_c$  (%) y factor de contaminación de magnitud  $\gamma_b$ .  $\theta_0 = 2$ . Primera fila ( $\gamma_b = 2$ ), segunda fila ( $\gamma_b = 5$ ).

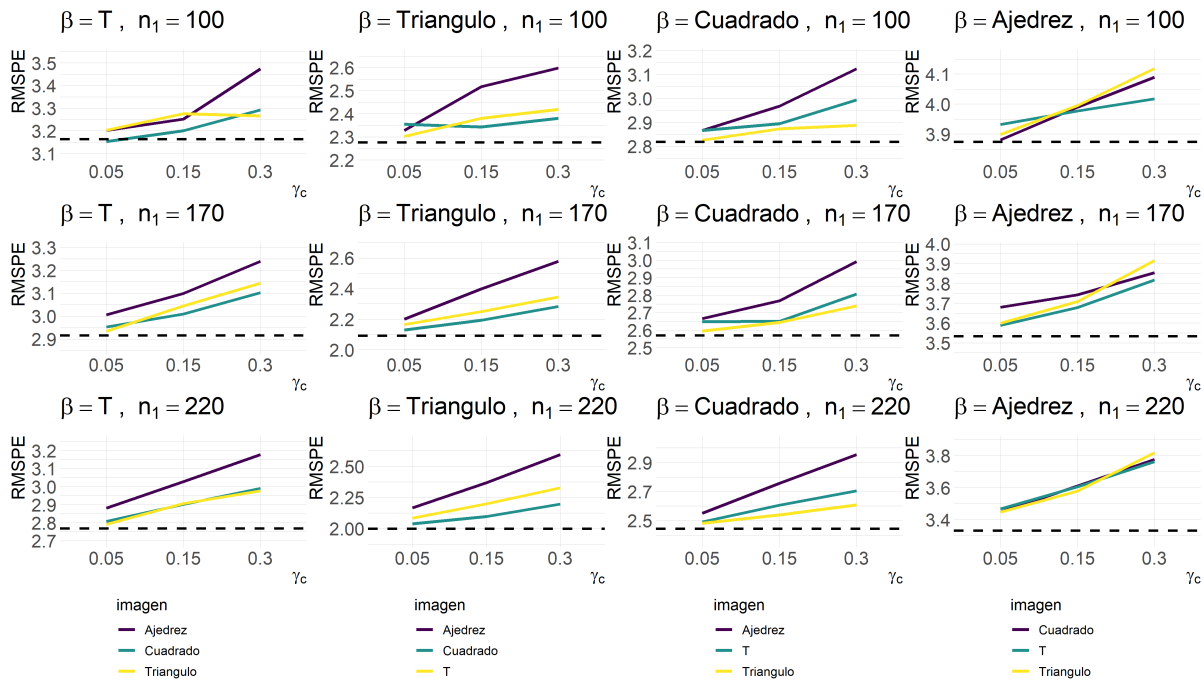
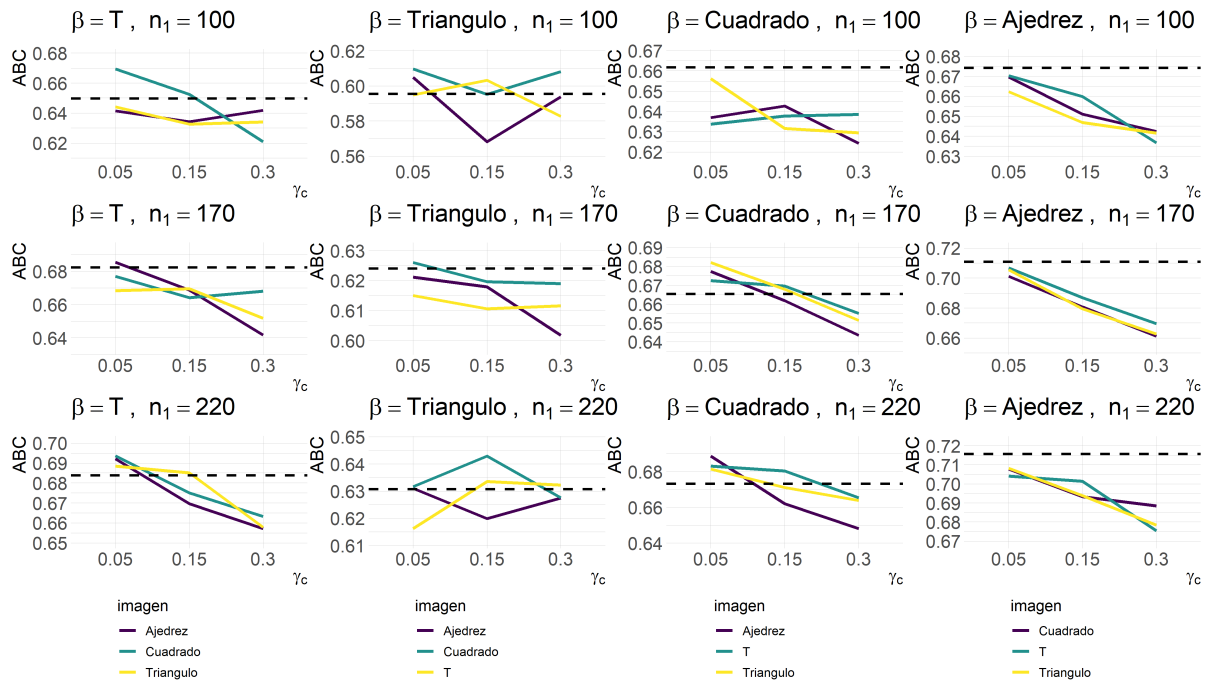
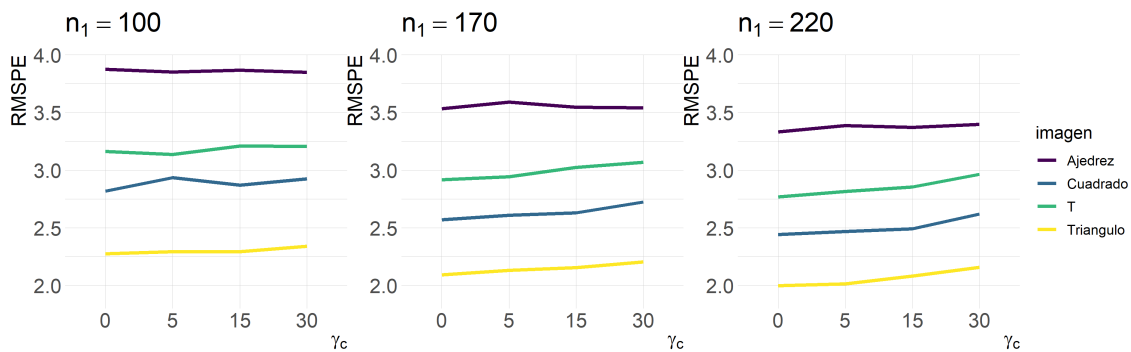


Figura 4-13.: Valores  $RMSPE$  promedio. Modelo de respuesta continua, bajo niveles de contaminación  $\gamma_c$  (%) y contaminación de forma Tipo 1.  $\theta_0 = 2$ . Primera columna ( $\beta_0 = T$ ), segunda columna ( $\beta_0 = Triángulo$ ), tercera columna ( $\beta_0 = Cuadrado$ ), cuarta columna ( $\beta_0 = Ajedrez$ ). Línea punteada indica valor  $RMSPE$  promedio bajo el escenario sin contaminación.





**Figura 4-14.:** Valores  $ABC$  promedio. Modelo de respuesta binaria, bajo niveles de contaminación  $\gamma_c$  (%) y contaminación de forma Tipo 1.  $\theta_0 = 2$ . Primera columna ( $\beta_0 = T$ ), segunda columna ( $\beta_0 = Triángulo$ ), tercera columna ( $\beta_0 = Cuadrado$ ), cuarta columna ( $\beta_0 = Ajedrez$ ). Línea punteada indica valor  $ABC$  promedio bajo el escenario sin contaminación.



**Figura 4-15.:** Valores  $RMSPE$  promedio. Modelo de respuesta continua, bajo niveles de contaminación  $\gamma_c$  (%) y contaminación de forma Tipo 2 (zoom del 50% en parámetro imagen).

En la Figura (4-15) se muestran los valores  $RMSPE$  promedio para distintos niveles de contaminación y bajo los distintos tamaños de muestra. Se observa que cuando la imagen parámetro tiene forma de T, Triángulo o Cuadrado, a medida que aumenta el nivel de contaminación peor es el nivel de predicción del modelo, aunque este detrimento no es abrupto.

En el caso del modelo con  $\beta_0 = \text{Ajedrez}$ , no se observan cambios evidentes en la capacidad de predicción del modelo, dado que el aumento de la imagen es casi idéntico a la imagen original como se puede ver en la Figura (4-2).

### 4.3. Identificación de observaciones influyentes en modelos GSIRM-TV

El objetivo de esta sección es determinar la efectividad de los métodos propuestos para la identificación de observaciones influyentes en los modelos escalar-imagen GSIRM-TV. En el estudio de simulación se comparará la aplicación del método basado en la medida *Directional Outlyingness* (DO), propuesta por [Rousseeuw and Hubert \(2018\)](#), la variación de este método basada en el uso de bootstrap para determinar el valor umbral del criterio con el cual se determina si una observación es influyente o no, y en el método *MS-PLOT* propuesto por [Dai and Genton \(2018\)](#).

En este segundo estudio de simulación se consideran los modelos escalar-imagen simples (sin intercepto):

$$Y_i = \langle X_i, \beta_0 \rangle + \epsilon_i \quad (4-12)$$

$$Y_i | X_i \sim \text{Bernoulli}(\mu_i)$$

$$\ln \left( \frac{\mu_i}{1 - \mu_i} \right) = \langle X_i, \beta_0 \rangle \quad (4-13)$$

Como en el caso anterior, el parámetro imagen del modelo original y del modelo contaminado provienen del conjunto de imágenes de la Figura (4-1). Para el proceso de simulación, análogo a la primera parte del estudio, se generan  $n_1$  observaciones para el conjunto de entrenamiento con  $n_c = \lceil \gamma_c n_1 \rceil$  observaciones del modelo contaminado,  $\gamma_c = \{5\%, 15\%, 30\%\}$ , y un conjunto de prueba de tamaño  $n_2 = \lceil 0.3n_1 \rceil$ . Se consideran tamaños de la muestra de entrenamiento de  $n_1 = 50, 70$  y se realizan  $R = 100$  replicaciones de cada uno de los escenarios. Cabe anotar que en esta segunda parte del estudio se definen tamaños de muestra menores a los de la primera parte, dado que para llevar a cabo los métodos propuestos es necesario estimar el parámetro imagen  $n_1$  veces en cada replicación. Lo cual tiene un costo computacional bastante alto.

Para los escenarios bajo contaminación de magnitud, se consideran factores de contaminación  $\gamma_b = 2, 5$ . En cada replicación de los escenarios, se estima el parámetro imagen omitiendo cada

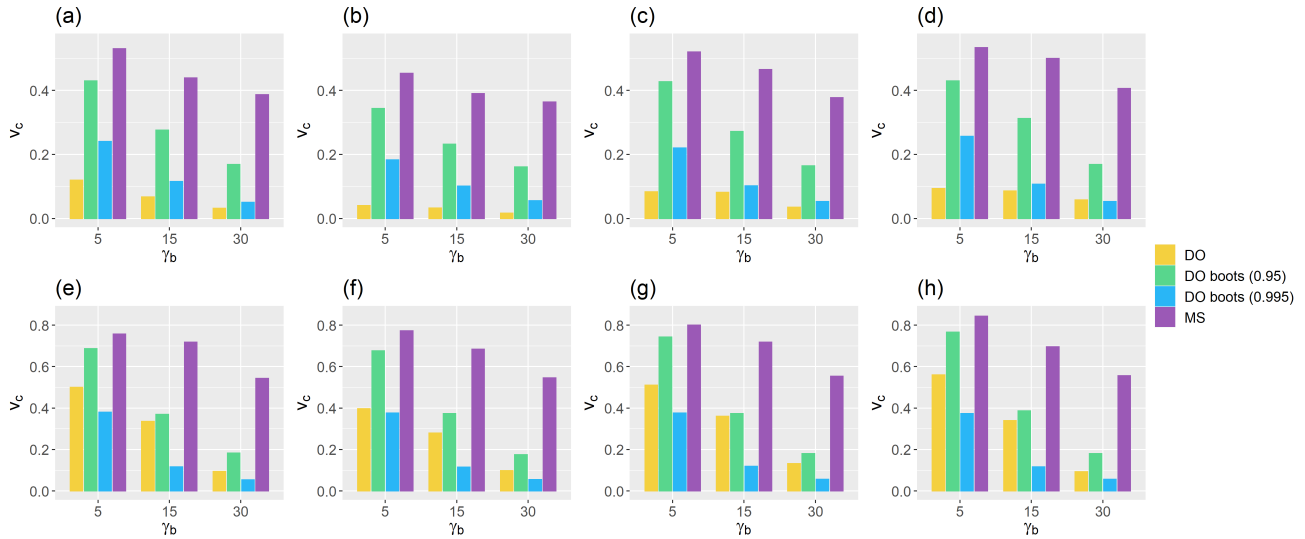
una de las  $n_1$  observaciones, y se llevan a cabo los métodos propuestos sobre el conjunto de  $n_1$  estimaciones, para identificar observaciones influyentes. Si dentro de la muestra generada se identifican estimaciones alejadas en el conjunto de  $n_1$  estimaciones, las observaciones son omitidas y el modelo se estima nuevamente. Así, se comparan los valores  $RMPSE$  (o  $ABC$ ) del modelo estimado originalmente y del modelo estimado sin las observaciones identificadas, para determinar el impacto de las observaciones influyentes en el desempeño del modelo en términos de predicción de la respuesta de las observaciones del conjunto de prueba.

En primer lugar se evalúa la capacidad de los métodos para identificar observaciones que provienen del modelo contaminado bajo contaminación de magnitud en el parámetro imagen. Para esto se definen las siguientes medidas: la tasa de correcta detección  $v_c$  (número de observaciones de modelo contaminado correctamente detectadas, dividido por el total de observaciones provenientes del modelo de contaminación) y la tasa de falsa detección  $f_c$  (número de observaciones de modelo original detectadas incorrectamente, dividido por el total de observaciones generadas del modelo original).

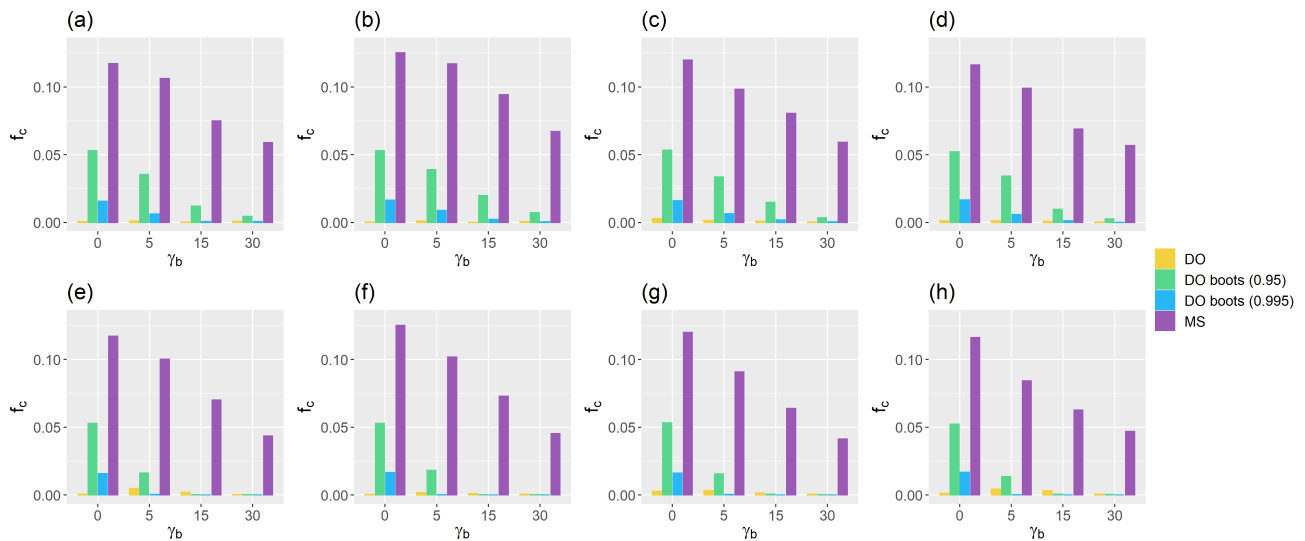
En la Figura (4-16) se muestran las tasas de correcta detección bajo los escenarios considerados. Se observa que en todos los escenarios el método con mejor desempeño es el MS-PLOT (en adelante MS), seguido de la propuesta basada en la medida DO con cuantil 0.95 calculado vía bootstrap (DO boost 0.95). Se evidencia que las tasas de detección son mayores cuando el nivel de contaminación es pequeño y disminuye a medida que aumenta el nivel de contaminación  $\gamma_c$ . Lo cual es de esperarse, dado que cuanto mayor es el nivel de contaminación, más difícil será la identificación dado el efecto enmascaramiento.

La Figura (4-17) muestra las tasas de falsa detección. Aunque el método MS tiene las tasas de correcta detección altas, también presenta una altas tasas de falsa detección, incluso en los escenarios en los que no hay contaminación en la muestra. Por el contrario, el método DO boots (0.95) mantiene tasas de falsa detección bajas, teniendo un desempeño mejor que los métodos DO y DO boots (0.995) en la detección de observaciones que provienen del modelo contaminado y que potencialmente pueden ser influyentes, teniendo en cuenta las afectaciones evidenciadas en el estudio de simulación de la sección anterior.

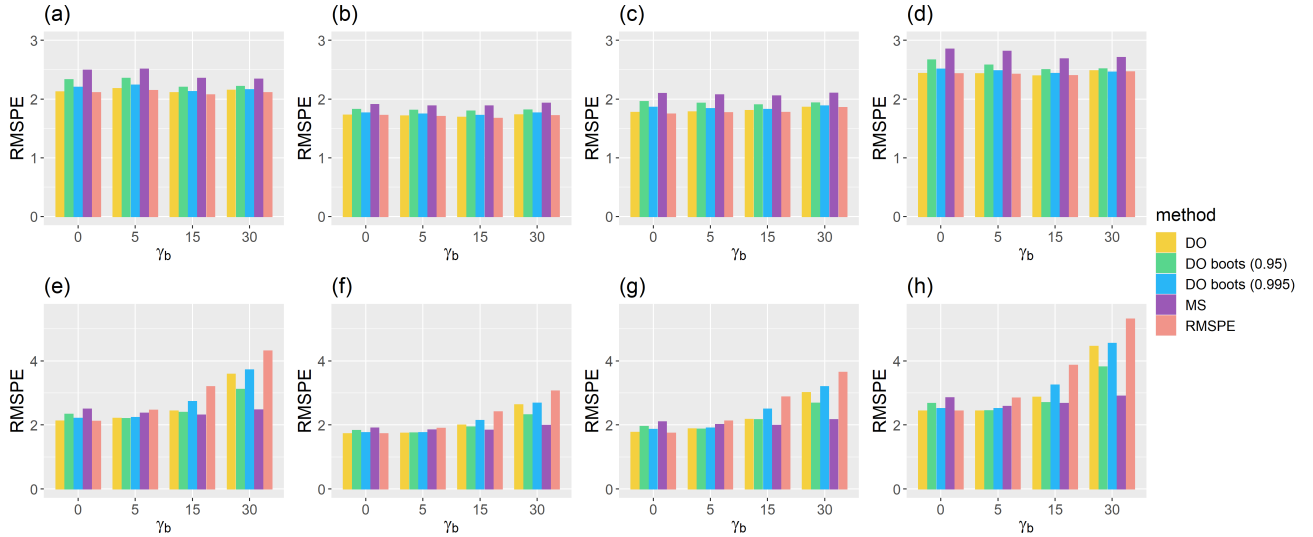
Ahora, si bien algunos de los métodos propuestos son efectivos en la detección de observaciones que provienen del modelo contaminado, es importante determinar si dichas observaciones tiene un impacto negativo en el desempeño del modelo en términos de predicción de nuevas observaciones. La Figura (4-18) muestra los valores  $RMSPE$  promedio de las replicaciones con muestras contaminadas y los valores  $RMPSE$  promedio obtenidos luego de omitir las observaciones identificadas por cada uno de los métodos.



**Figura 4-16.:** Tasas de correcta detección ( $v_c$ ) promedio en 100 replicaciones. Modelo de respuesta continua sin intercepto, bajo niveles de contaminación  $\gamma_c$  (%) y factor de contaminación de magnitud  $\gamma_b$ . Primera fila ( $\gamma_b = 2$ ), segunda fila ( $\gamma_b = 5$ ).  $n = 70$ . Gráficas (a) y (e):  $\beta_0 = T$ , (b) y (f):  $\beta_0 = \text{Triángulo}$ , (c) y (g):  $\beta_0 = \text{Cuadrado}$ , (d) y (h):  $\beta_0 = \text{Ajedrez}$ .



**Figura 4-17.:** Tasas de falsa detección ( $f_c$ ) promedio en 100 replicaciones. Modelo de respuesta continua sin intercepto, bajo niveles de contaminación  $\gamma_c$  (%) y factor de contaminación de magnitud  $\gamma_b$ . Primera fila ( $\gamma_b = 2$ ), segunda fila ( $\gamma_b = 5$ ). Gráficas (a) y (e):  $\beta_0 = T$ , (b) y (f):  $\beta_0 = \text{Triángulo}$ , (c) y (g):  $\beta_0 = \text{Cuadrado}$ , (d) y (h):  $\beta_0 = \text{Ajedrez}$ .



**Figura 4-18.:** Valores  $RMSPE$  promedio en 100 repeticiones y valores  $RMSPE$  obtenidos al omitir las observaciones identificadas como influyentes por cada método. Modelo de respuesta continua sin intercepto, bajo niveles de contaminación  $\gamma_c$  (%) y factor de contaminación de magnitud  $\gamma_b$ . Primera fila ( $\gamma_b = 2$ ), segunda fila ( $\gamma_b = 5$ ). Gráficas (a) y (e):  $\beta_0 = T$ , (b) y (f):  $\beta_0 = \text{Triángulo}$ , (c) y (g):  $\beta_0 = \text{Cuadrado}$ , (d) y (h):  $\beta_0 = \text{Ajedrez}$ .

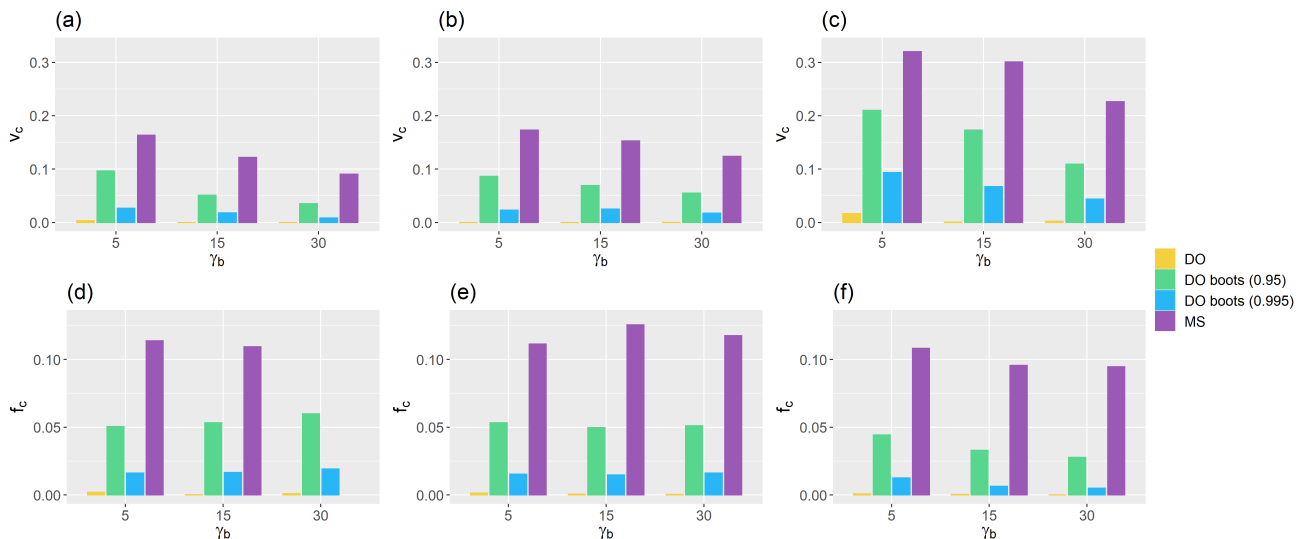
Se observa que cuando el factor de contaminación es  $\gamma_b = 2$ , los valores  $RMPSE$  obtenidos al omitir las observaciones identificadas como posibles influyentes son similares a los valores obtenidos con todas las observaciones de la muestra de entrenamiento. Lo cual es esperado teniendo en cuenta que bajo ese factor de contaminación de magnitud, no se afectan gravemente las estimaciones de los parámetros y en consecuencia, la predicción, como se mostró en la primera parte.

Por otro lado, bajo un factor de contaminación  $\gamma_b = 5$ , se observa una mejoría en la predicción del modelo al omitir las observaciones identificadas con cualquiera de los métodos, especialmente cuando el nivel de contaminación es mayor al 5%. Cabe anotar que cuando el nivel de contaminación aumenta, los mejores resultados en términos de predicción se obtienen al omitir las observaciones identificadas con el método MS, seguido del método DO boost 0.95.

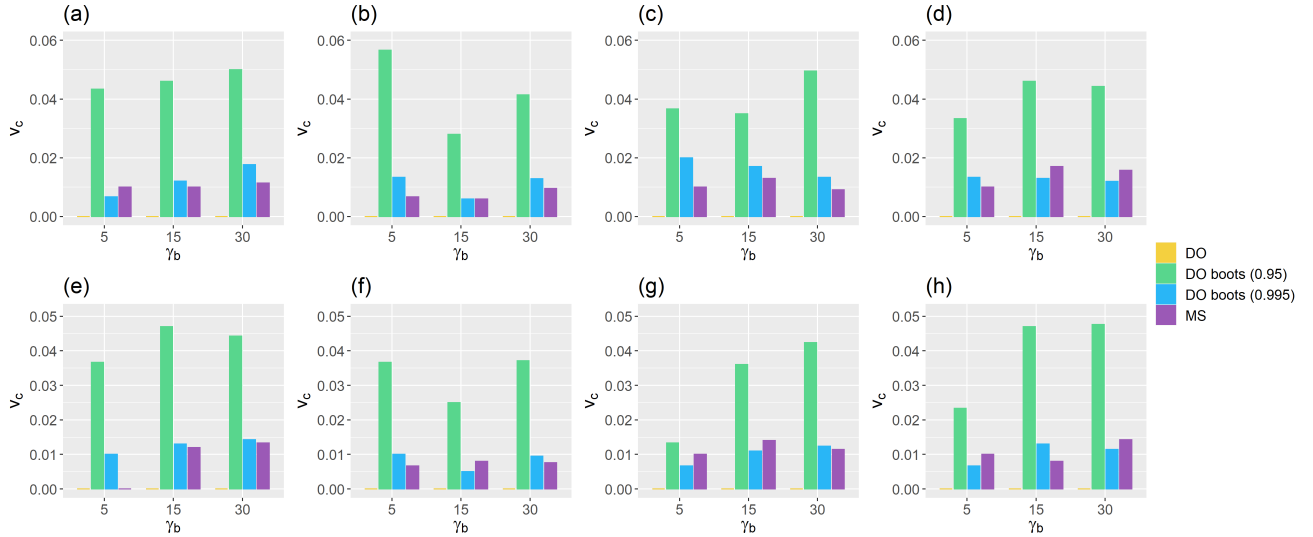
Por último, se presentan las tasas de correcta detección y de falsa detección bajo cambios de forma Tipo 1 cuando el parámetro imagen original es  $\beta_0 = T$ , (Figura (4-19)). Como se mencionó anteriormente este tipo de contaminación no afecta de forma notoria las estimaciones del parámetro imagen. Se observa que las tasas de correcta detección son bajas en comparación con las observadas bajo contaminación de magnitud. Nuevamente, el método

MS presenta las tasas de detección más altas. Sin embargo, también presenta altas tasas de falsa detección. El mismo comportamiento se observa cuando  $\beta_0$  tiene forma de Triángulo, Cuadrado o Ajedrez, como se puede verificar en las Tablas (A-20) y (A-21) del Anexo (A).

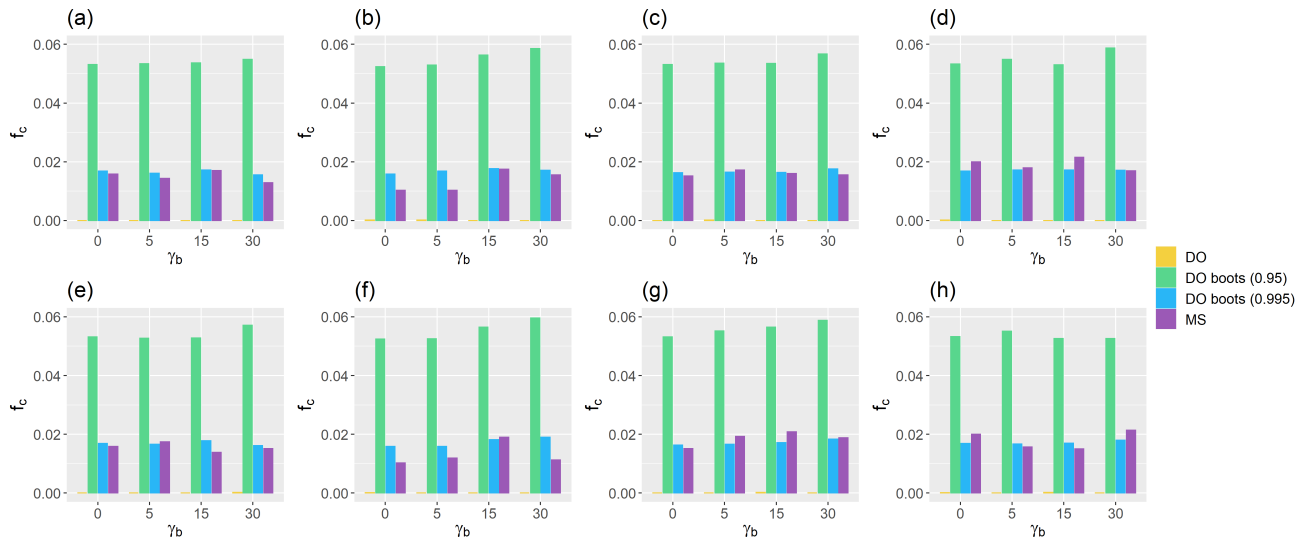
En el caso del modelo de respuesta binaria, en las Figuras (4-20) y (4-21) se muestran los promedios de las tasas de correcta detección y falsa detección, respectivamente. En todos los escenarios simulados bajo este modelo los métodos propuestos presentan bajas tasas de correcta detección. Lo cual es esperado, teniendo en cuenta los resultados de la primera parte del estudio de simulación, donde se observó que la contaminación de magnitud en el parámetro imagen, no deteriora fuertemente las estimaciones del mismo. Adicionalmente, en la Figura (4-22) se muestran las tasas de correcta y falsa detección bajo contaminación de cambio de forma - Tipo 1 en el modelo de respuesta binaria con parámetro imagen  $\beta_0 = T$ . En los escenarios considerados bajo este modelo, las tasa de correcta detección no superan el 15%, lo que es consecuente con lo observado en el estudio de simulación de la sección anterior. Los resultados bajo un tamaño de muestra  $n = 50$ , y distintas formas de la imagen  $\beta_0$ , son similares como se observa en las Tablas (A-18), (A-19), (A-22) y (A-23) del Anexo A.



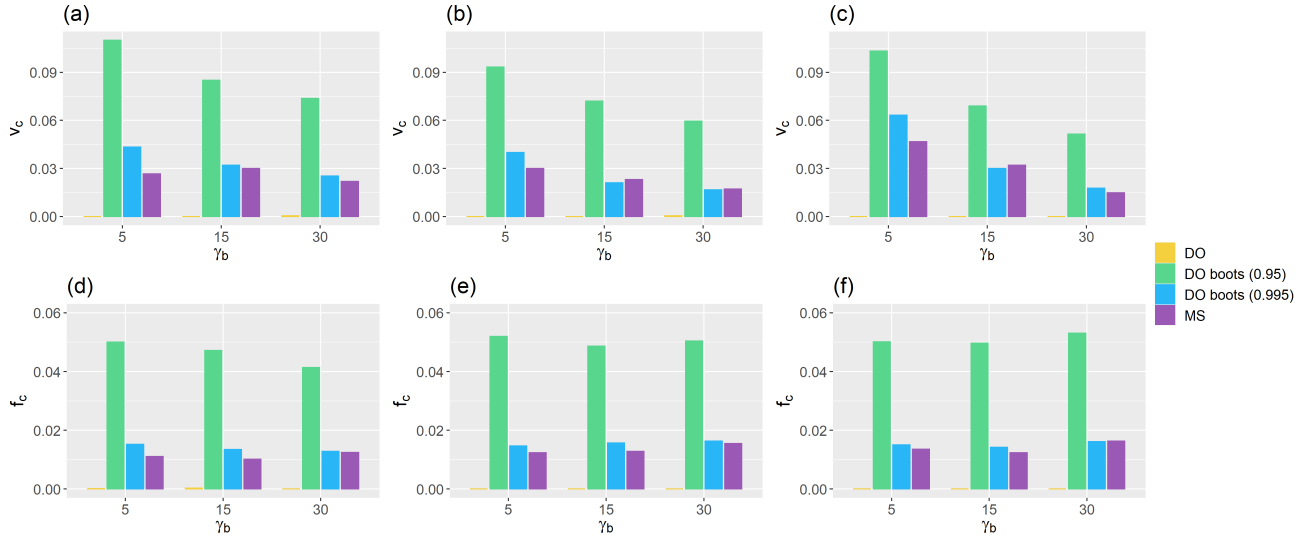
**Figura 4-19.:** Primera fila: Tasas de correcta detección ( $v_c$ ) promedio. Segunda fila: Tasas de falsa detección ( $f_c$ ) promedio. Modelo de respuesta continua sin intercepto con  $\beta_0 = T$ , bajo niveles de contaminación  $\gamma_c$  (%) y contaminación de forma Tipo 1. Gráficas (a) y (d):  $\beta_1 = \text{Triángulo}$ , (b) y (e):  $\beta_1 = \text{Cuadrado}$ , (c) y (f):  $\beta_1 = \text{Ajedrez}$ .



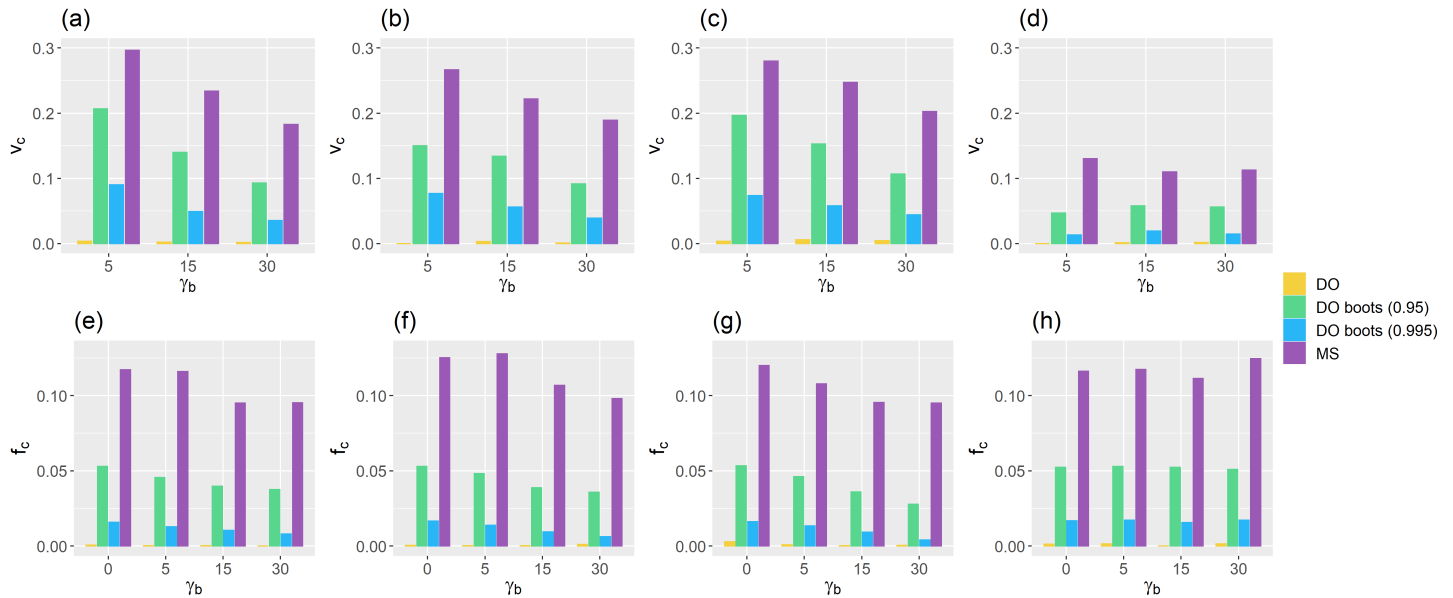
**Figura 4-20.:** Tasas de correcta detección ( $v_c$ ) promedio en 100 repeticiones. Modelo de respuesta binaria sin intercepto, bajo niveles de contaminación  $\gamma_c$  (%) y factor de contaminación de magnitud  $\gamma_b$ .  $n = 70$ . Primera fila ( $\gamma_b = 2$ ), segunda fila ( $\gamma_b = 5$ ). Gráficas (a) y (e):  $\beta_0 = T$ , (b) y (f):  $\beta_0 = \text{Triángulo}$ , (c) y (g):  $\beta_0 = \text{Cuadrado}$ , (d) y (h):  $\beta_0 = \text{Ajedrez}$ .



**Figura 4-21.:** Tasas de falsa detección ( $f_c$ ) promedio en 100 repeticiones. Modelo de respuesta binaria sin intercepto, bajo niveles de contaminación  $\gamma_c$  (%) y factor de contaminación de magnitud  $\gamma_b$ .  $n = 70$ . Primera fila ( $\gamma_b = 2$ ), segunda fila ( $\gamma_b = 5$ ). Gráficas (a) y (e):  $\beta_0 = T$ , (b) y (f):  $\beta_0 = \text{Triángulo}$ , (c) y (g):  $\beta_0 = \text{Cuadrado}$ , (d) y (h):  $\beta_0 = \text{Ajedrez}$ .



**Figura 4-22.:** Primera fila: Tasas de correcta detección ( $v_c$ ) promedio. Segunda fila: Tasas de falsa detección ( $f_c$ ) promedio. Modelo de respuesta binaria sin intercepto con  $\beta_0 = T$ , bajo niveles de contaminación  $\gamma_c(\%)$  y contaminación de forma Tipo 1. Gráficas (a) y (d):  $\beta_1 = \text{Triángulo}$ , (b) y (e):  $\beta_1 = \text{Cuadrado}$ , (c) y (f):  $\beta_1 = \text{Ajedrez}$ .



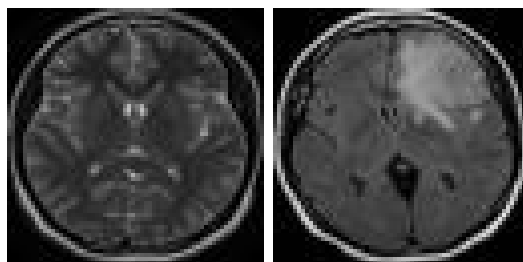
**Figura 4-23.:** Primera fila: Tasas de correcta detección ( $v_c$ ) promedio. Segunda fila: Tasas de falsa detección ( $f_c$ ) promedio. Modelo de respuesta continua sin intercepto con  $\beta_0 = T$ , bajo niveles de contaminación  $\gamma_c(\%)$  y contaminación de forma Tipo 2 (zoom del 50% en parámetro imagen). Gráficas (a) y (e):  $\beta_0 = T$ , (b) y (f):  $\beta_0 = \text{Triángulo}$ , (c) y (g):  $\beta_0 = \text{Cuadrado}$ , (d) y (h):  $\beta_0 = \text{Ajedrez}$ .



Por último, En cuanto a la capacidad de los métodos propuestos para la detección de observaciones influyentes bajo la contaminación de forma Tipo 2, los resultados son similares a los observados bajo la contaminación de cambio de forma Tipo 1 considerada anteriormente. En general, el método MS tiene las tasas más altas de correcta detección y falsa detección. Mientras que el método DO boost (0.95) muestra tasas de correcta detección superiores a los dos métodos restantes, manteniendo tasas de falsa detección bajas (Figura (4-23)). Aunque ningún método muestra alta efectividad en la detección de observaciones del modelo contaminado, teniendo en cuenta que este tipo de contaminación no genera un detrimento significativo de las estimaciones del modelo, como se puede observar en las Tablas (A-13), (A-14) y (A-15) del Anexo A.

## 4.4. Aplicación en datos reales

Para ilustrar los métodos propuestos para la identificación de observaciones influyentes, se estima un modelo de respuesta binaria con función de enlace *logit*, a partir de un conjunto de imágenes de resonancia magnética (MRI). El conjunto de datos consta de 253 imágenes MRI del cerebro de 98 personas sanas y 155 personas con tumor cerebral. El conjunto de datos puede ser descargado de (<https://www.kaggle.com/datasets/navoneel/brain-mri-images-for-brain-tumor-detection>).



**Figura 4-24.:** Ejemplo MRI de paciente sano (izquierda) y paciente con tumor cerebral (derecha).

La Figura (4-24) muestra un ejemplo de la imagen asociada a una persona sana y la imagen tomada en una persona con tumor cerebral. Para estimar el modelo y determinar su efectividad en la identificación de pacientes con algún tipo de tumor, se divide el conjunto de datos en un conjunto de entrenamiento y uno de prueba. El conjunto de entrenamiento está conformado por 78 imágenes de pacientes sanos y 124 imágenes de pacientes enfermos, seleccionados aleatoriamente. Por otra parte, el conjunto de prueba queda conformado por las imágenes de 20 pacientes sanos y 31 pacientes enfermos. A continuación, se presenta el

modelo lineal generalizado a estimar a partir del conjunto de datos de entrenamiento:

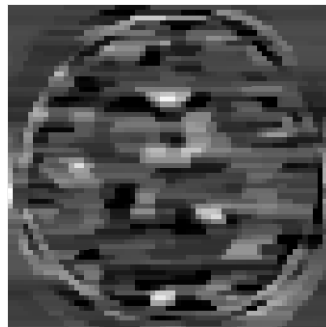
$$Y_i|X_i \sim \text{Bernoulli}(\mu_i)$$

$$\ln\left(\frac{\mu_i}{1-\mu_i}\right) = \langle X_i, \beta_0 \rangle + \theta_0, \quad (4-14)$$

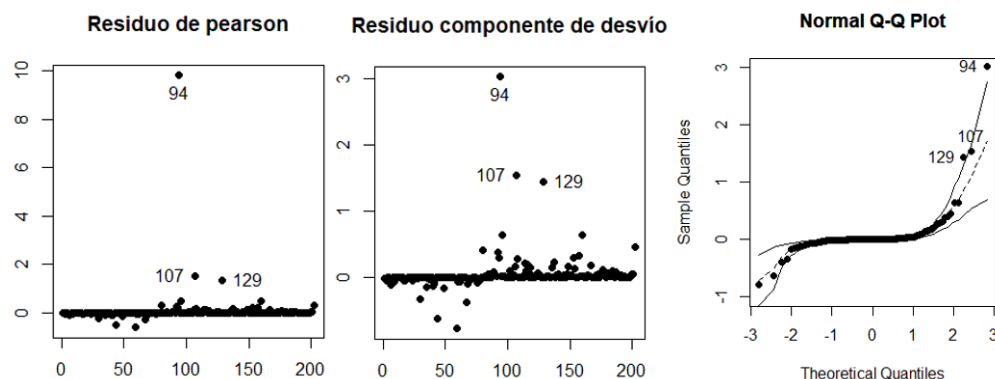
con  $i = 1, \dots, 202$  y  $Y_1, \dots, Y_{202}$ , variables aleatorias independientes. La covariable  $X_i$  corresponde a la imagen MRI del paciente  $i$ -ésimo. La variable respuesta toma un valor igual a 1 si el paciente presenta tumor cerebral y 0 en otro caso. En la Figura (4-25) se muestra la estimación del parámetro imagen  $\beta_0$ , obtenida a partir de los datos del conjunto de entrenamiento. Por otra parte, la estimación del intercepto es  $\hat{\theta}_0 = -0.5488$ .

Para evaluar el ajuste del modelo, se calculan los residuos componente de desvío y residuos de pearson. En la Figura (4-26) se presentan las gráficas de dichos residuos, junto con el gráfico de las bandas del 95 % de confianza para los residuos componente de desvío. La Figura muestra que los residuos presentan valores pequeños en su mayoría, y el gráfico de las bandas de confianza muestra que en general los residuos se ubican dentro de las mismas. Lo cual indica que el modelo propuesto parece ser adecuado. Sin embargo, se señalan residuos de 3 unidades muestrales que presentan residuos grandes. Estas observaciones son la 107 y las observaciones 94 y 129, cuyos residuos se salen de las bandas de confianza.

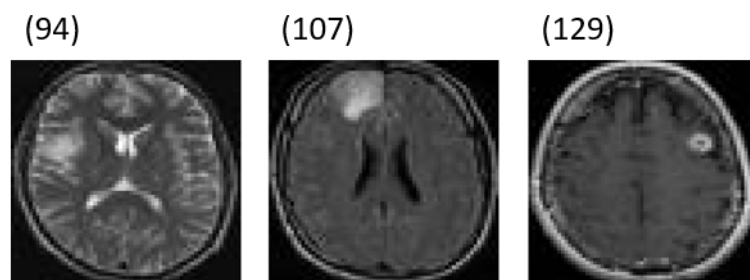
Analizando estas unidades muestrales, se evidencia que corresponden a pacientes con tumor cerebral. Por lo tanto, las respuestas asociadas son  $y_{94} = 1$ ,  $y_{107} = 1$  y  $y_{129} = 1$ . Por otro lado, las probabilidades estimadas a partir del modelo de que correspondan a pacientes con tumor, son respectivamente  $\hat{\mu}_{94} = 0.01$ ,  $\hat{\mu}_{107} = 0.30$  y  $\hat{\mu}_{129} = 0.35$ . En la Figura (4-27) se muestran las imágenes MRI asociadas a estos individuos. En todos los casos, se observa que el tumor cerebral es difícil de identificar a simple vista, bien sea por el tamaño del tumor o por la intensidad de la imagen en la región asociada. Lo cual genera una difícil identificación del tipo de paciente por parte del modelo, y por tanto, residuos grandes.



**Figura 4-25.:** Estimación del parámetro  $\beta$ .



**Figura 4-26.:** Residuos de Pearson (Izquierda), Residuos componente de desvío (Centro) y Bandas del 95 % de confianza de los residuos componente de desvío (Derecha).



**Figura 4-27.:** Imágenes MRI de unidades muestrales 94, 107 y 129 de la muestra de entrenamiento.

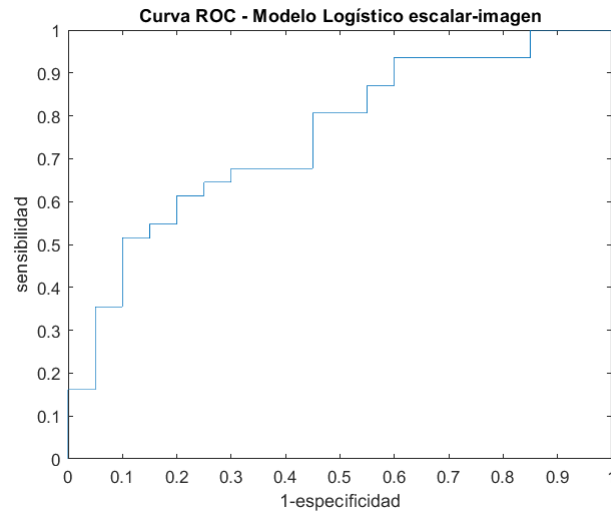
Por otro lado, la Tabla (4-2) presenta la matriz de confusión asociada a los datos de prueba, con lo cual se determina que la precisión del modelo es de 0.6863. Adicionalmente, la Figura (4-28) muestra la curva ROC asociada, con la cual se tiene un valor  $ABC = 0.7516$ , el cual indica que el modelo predice de forma aceptable la condición de un paciente (sano o enfermo), a través de la imagen MRI de su cerebro.

Como se mencionó en el capítulo anterior, es este trabajo se propone medir la influencia de las observaciones en la estimación del parámetro imagen  $\beta$ , aplicando los métodos propuestos por [Rousseeuw and Hubert \(2018\)](#) y [Dai and Genton \(2018\)](#), en adelante DO y MS respectivamente, sobre el las 202 estimaciones obtenidas al omitir cada una de las observaciones del conjunto de entrenamiento. Adicionalmente, se implementan las variantes del método DO, calculando el valor umbral contra el cual se comparan los valores  $Z_{DO_i}$ ,  $i = 1, \dots, 202$ , de la expresión (3-21), estimando el cuantil 0.955 ó 0.95 de la estadística  $Z_{DO}$ , vía bootstrap. En adelante, se denotan a estas variaciones del método DO como DO boost(0.955) y DO boost(0.95), dependiendo del cuantil estimado. En la Figura (4-29) se presenta el *functional outlier map* (FOM) y el MS-PLOT resultantes de la aplicación de los métodos DO y MS

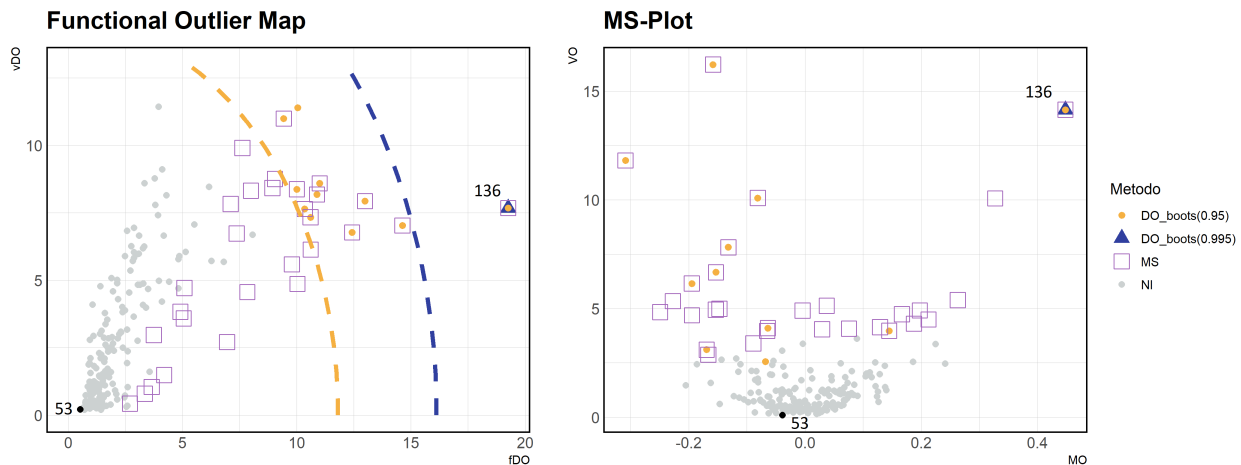
sobre el conjunto de estimaciones  $\{\hat{\beta}_{(1)}, \dots, \hat{\beta}_{(202)}\}$ , siendo  $\hat{\beta}_{(i)}$  la estimación de la imagen  $\beta$  obtenida al omitir la  $i$ -ésima observación del conjunto de entrenamiento.

		Valor predicho	
		1	0
Valor verdadero	1	14	6
	0	10	21

**Tabla 4-2.:** Matriz de confusión de datos de prueba.



**Figura 4-28.:** Curva ROC asociada al modelo estimado.



**Figura 4-29.:** FOM y MS-PLOT de estimaciones de  $\hat{\beta}_{(i)}$ ,  $i = 1, \dots, 202$ .

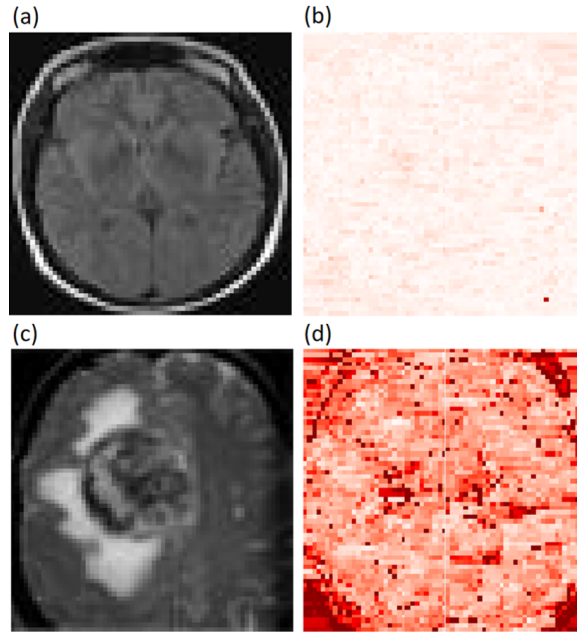
En color gris se marcan los puntos correspondientes a las observaciones que no fueron identificadas como influyentes por ninguno de los 4 métodos (NI). En color amarillo se marcan los puntos de las observaciones identificadas como influyentes por el método DO boost(0.95), los cuadrados de color morado marcan las identificadas por el método DO boost(0.995) y los triángulos de color azul, las identificadas por el método MS. Las líneas punteadas de color amarillo y azul que se observan en el FOM, corresponden a las elipses definidas por los umbrales obtenidos a partir de los métodos DO boost(0.95) y DO boost(0.995) respectivamente. Cabe anotar que a través del método DO, no se identificó ninguna observación como influyente. Adicionalmente, en color negro se marca la observación 53, que corresponde a una de las observaciones con menor influencia en la estimación del parámetro imagen  $\beta$ , de acuerdo con su ubicación en el FOM y en el MS-PLOT. Por último, en la Figura se observa que la unidad muestral 136 fue identificada como influyente por todos los métodos, a excepción del método DO, que no identificó ninguna observación como tal.

En la Figura (4-30) se muestran las imágenes MRI de la unidad muestral 53 (Figura (4-30)(a)) y la unidad muestral 136 (Figura (4-30)(c)), la cual fue identificada como una observación influyente por 3 de los 4 métodos. Los mapas de calor de los valores  $DO$  pixel a pixel propuestos por [Rousseeuw and Hubert \(2018\)](#) asociados a las estimaciones  $\hat{\beta}_{(53)}$  y  $\hat{\beta}_{(136)}$ , se muestran en las Figuras (4-30)(b) y (4-30)(d) respectivamente. Estos mapas de calor, permiten identificar las áreas de las estimaciones que difieren del conjunto general de estimaciones  $\{\hat{\beta}_{(1)}, \dots, \hat{\beta}_{(202)}\}$ .

Dado que la estimación  $\hat{\beta}_{(53)}$  presenta valores bajos de las medidas  $fDO$ ,  $vDO$ ,  $MO$  y  $VO$ , como se observa en la Figura (4-29), se puede decir que la unidad muestral 53 es poco influyente en la estimación del parámetro imagen  $\beta$  y en consecuencia el mapa de calor asociado no presenta regiones de color oscuro. Por otra parte, la estimación  $\hat{\beta}_{(136)}$  presenta valores altos en las medidas  $fDO$ ,  $vDO$ ,  $MO$  y  $VO$ , por lo cual fue identificada como influyente. Al observar el mapa de calor, se identifican áreas oscuras donde la estimación  $\hat{\beta}_{(136)}$  difiere considerablemente de las demás estimaciones. Dichas áreas se ubican en las esquinas de la imagen, principalmente. Analizando la imagen MRI de la unidad muestral 153, se observa que el área del cerebro no se encuentra bien definida dada una baja resolución de la imagen, además de una baja luminosidad. Lo cual podría ser una causa de que esta unidad muestral sea identificada como influyente, contrario a la unidad muestral 53, cuya imagen MRI se encuentra bien definida.

Por último, para analizar el impacto de la observación 153 en la predicción de la variable respuesta de las unidades del conjunto de prueba, se estima el modelo omitiendo la observación 153 del conjunto de entrenamiento. En la Tabla (4-3) se muestra la matriz de confusión asociada al nuevo modelo estimado. Comparándola con la matriz presentada en la Tabla (4-2), se evidencia que una de las observaciones mal clasificada como una persona con tumor

cerebral, por el modelo estimado con todas las observaciones del conjunto de prueba, ahora es clasificada correctamente como una persona sana, por el modelo estimado omitiendo la observación identificada como influyente. Esto resulta en una precisión del modelo de 0.7059 y un valor  $ABC = 0.7548$ . Lo cual muestra que el modelo estimado omitiendo la observación 153, tiene un mejor desempeño en términos de predicción.



**Figura 4-30.:** Imágenes MRI de unidades muestrales 53 (a) y 136 (c) y mapas de calor de valores DO de estimaciones  $\hat{\beta}_{(53)}$  (b) y  $\hat{\beta}_{(136)}$  (d).

		Valor predicho	
		1	0
Valor verdadero	1	14	6
	0	9	22

**Tabla 4-3.:** Matriz de confusión de datos de prueba, omitiendo la unidad muestral 136 de conjunto de entrenamiento.

# 5. Conclusiones y recomendaciones

## 5.1. Conclusiones

Los modelos escalar-imagen constituyen una poderosa herramienta para el análisis de imágenes vistas como información auxiliar para predecir variables escalares, como por ejemplo, condiciones clínicas de pacientes con algún tipo de patología. En particular, los modelos GSIRM-TV se destacan dentro de las propuestas de modelos escalar-imagen, puesto que suponen que el parámetro imagen  $\beta$  pertenece al espacio de variación acotada  $BV(\Omega)$ . Lo cual es una ventaja en comparación con métodos desarrollados sobre el espacio de funciones cuadrado integrables  $L_2(\Omega)$ , puesto que el espacio  $BV(\Omega)$  contiene distintas funciones con saltos y bordes, lo cual caracteriza a muchos tipos de imágenes como las imágenes MRI. Esto permite que los métodos de regularización basados en este espacio capturen mejor la información de este tipo de imágenes, contrario a los métodos basados en espacio de funciones suaves como  $L_2(\Omega)$ .

Dado su reciente desarrollo y complejidad de implementación, aún existen vacíos en la literatura para un análisis completo de este tipo de modelos, incluyendo el diagnóstico de los mismos. En particular, en la literatura no se ha abordado el análisis de influencia en modelos escalar-imagen, el cual es una parte fundamental en el análisis de cualquier tipo de modelo estadístico, puesto que permite identificar particularidades dentro de las unidades de análisis que puede revelar información novedosa de las mismas. En este trabajo se evidencia que este tipo de modelos puede verse afectado ante la presencia de observaciones que provienen de otra población.

La primera parte del estudio de simulación muestra que las estimaciones del modelo tanto del parámetro imagen como la del intercepto, pueden verse afectadas por la presencia de observaciones que provienen de modelos con cambio en el parámetro imagen en magnitud o en forma. Si bien el deterioro de las estimaciones se da principalmente en el modelo de respuesta continua, el modelo de respuesta binaria con función de enlace *logit*, también es susceptible de presentar afectación en la estimación del intercepto, principalmente. Aunque bajo ciertas condiciones también puede verse afectada la estimación del parámetro imagen. Este detrimento puede ser alto cuando aumenta el nivel de contaminación en la muestra, lo cual es un indicativo de que los modelos lineales generalizados escalar-imagen estimados vía variación total pueden no ser robustos a observaciones atípicas.

Adicionalmente, se proponen métodos para la identificación de observaciones influyentes en la estimación del parámetro imagen, siguiendo la filosofía de la distancia de Cook en modelos lineales generalizados clásicos. Estos métodos se basan en las nociones de *outlyingness* en datos funcionales, y específicamente, en los procedimientos del *functional outlier map* (FOM) propuesto por [Rousseeuw and Hubert \(2018\)](#), y el MS-PLOT propuesto por [Dai and Genton \(2018\)](#), para la identificación de *outliers* en conjunto de datos funcionales univariados y multivariados.

La segunda parte del estudio de simulación, muestra que los métodos propuestos pueden tener una alta efectividad en la detección de observaciones con características particulares que pueden afectar la estimación del parámetro imagen y en consecuencia, el desempeño del modelo en términos de predicción de nuevas observaciones. Los métodos más efectivos según los resultados del estudio, son el método basado en el MS-PLOT, seguido del método basado en *Directional Outlyingness* con cálculo de cuantil 0.95, calculado vía bootstrap. Aunque el método MS-PLOT sufre de una alta tasa de falsa detección, por lo que puede ser demasiado sensible para realizar la labor de detección de observaciones influyentes.

Finalmente, la aplicación en el conjunto de datos reales ilustra el uso de los métodos propuestos para la identificación de observaciones influyentes en los modelos GSIRM-TV. El uso de estos métodos, en conjunto con métodos gráficos como los mapas de calor de los valores *DO*, constituyen una herramienta útil y efectiva para identificar particularidades dentro de las unidades muestrales que puedan afectar las estimaciones del modelo y en consecuencia, su desempeño en términos de predicción de nuevas observaciones.

## 5.2. Recomendaciones

Dados los resultados del estudio de simulación y los resultados del ejemplo con datos reales, se evidencia que el método MS-PLOT para la identificación de observaciones influyentes puede ser bastante sensible. Por lo cual se recomienda el método basado en *Directional Outlyingness* con cálculo de cuantil 0.95, calculado vía bootstrap (o en su defecto el mismo método con uso del cuantil 0.995), puesto que el estudio de simulación muestra que este procedimiento tiene buen desempeño en la detección de observaciones que provienen de otra población, manteniendo tasas de falsa detección bajas.

Cabe anotar que este estudio de simulación muestra que los modelos lineales generalizados escalar-función estimados vía variación total no son robustos ante la presencia de observaciones atípicas. Sin embargo, es necesario realizar la validación de forma teórica para determinar, por ejemplo, el punto de quiebre de las estimaciones de estos modelos, así como la función de influencia asociada a los estimadores.



Por último, es importante resaltar que el estudio de simulación de este trabajo abarcó un número limitado de escenarios, dado el alto costo computacional que involucra la estimación de este tipo de modelos, incluso considerando imágenes de baja resolución. Como trabajo futuro, es importante evaluar la robustez de estos modelos bajo otros tipos de cambios en el parámetro imagen, como por ejemplo rotaciones de las imágenes, o cambios en los parámetros escalares. También se hace necesario proponer un método para la identificación de observaciones influyentes que considere la afectación tanto del parámetro imagen como de los parámetros escalares.

## **A. Anexo: Resultados Estudio de Simulación**

$\beta_0$	$\gamma_c$ (%)	$\gamma_b$	$RMSPE$	$d(\beta, \hat{\beta})$	$d_2(\beta_v, \hat{\beta}_v)$	$d(\theta, \hat{\theta})$	$d_2(\theta, \hat{\theta})$	$\hat{\theta}$	Tiempo (Seg)
T	0	1.0	3.20 (0.40)	1.18 ( 65.09)	30.39 (0.75)	0.32 (0.22)	0.15 (0.19)	1.76	1231.77
T	5	0.2	3.17 (0.41)	1.19 ( 61.64)	30.24 (0.84)	0.32 (0.22)	0.15 (0.21)	1.77	997.47
T	5	0.5	3.19 (0.43)	1.28 ( 93.07)	30.21 (0.89)	0.35 (0.25)	0.18 (0.24)	1.76	984.50
T	5	2.0	3.60 (0.54)	1.91 (403.18)	34.20 (2.96)	0.47 (0.36)	0.35 (0.51)	1.70	996.66
T	5	5.0	3.23 (0.43)	1.16 ( 64.16)	30.81 (0.81)	0.29 (0.22)	0.13 (0.19)	1.80	999.91
T	15	0.2	3.17 (0.43)	1.16 ( 63.52)	30.40 (0.78)	0.27 (0.20)	0.12 (0.15)	1.80	1015.47
T	15	0.5	3.19 (0.43)	1.44 (123.87)	30.57 (0.81)	0.38 (0.25)	0.20 (0.23)	1.75	1020.33
T	15	2.0	4.33 (0.65)	2.75 (457.99)	41.42 (4.33)	0.67 (0.52)	0.72 (1.02)	1.58	1024.22
T	15	5.0	3.25 (0.45)	1.12 ( 60.64)	31.40 (0.68)	0.31 (0.22)	0.14 (0.18)	1.78	1042.64
T	30	0.2	3.21 (0.47)	1.12 ( 62.56)	30.65 (0.78)	0.30 (0.21)	0.13 (0.18)	1.76	1063.62
T	30	0.5	3.28 (0.40)	1.60 (135.15)	30.96 (0.85)	0.39 (0.29)	0.24 (0.32)	1.71	1065.50
T	30	2.0	5.03 (0.76)	3.48 (442.92)	48.83 (4.89)	0.74 (0.57)	0.87 (1.30)	1.65	1063.91
T	30	5.0	3.16 (0.42)	1.19 ( 65.39)	30.14 (0.85)	0.34 (0.24)	0.18 (0.22)	1.76	972.75
Triangulo	0	1.0	2.34 (0.29)	0.91 ( 44.60)	20.83 (0.62)	0.23 (0.16)	0.08 (0.09)	1.83	1363.09
Triangulo	5	0.2	2.30 (0.28)	0.91 ( 44.56)	20.66 (0.60)	0.23 (0.15)	0.08 (0.09)	1.82	1248.08
Triangulo	5	0.5	2.30 (0.32)	0.96 ( 60.54)	20.86 (0.61)	0.24 (0.17)	0.09 (0.11)	1.81	989.48
Triangulo	5	2.0	2.65 (0.40)	1.37 (257.42)	23.78 (2.18)	0.34 (0.25)	0.18 (0.24)	1.76	994.39
Triangulo	5	5.0	2.33 (0.32)	0.90 ( 46.47)	21.21 (0.60)	0.24 (0.17)	0.09 (0.11)	1.82	993.20
Triangulo	15	0.2	2.33 (0.29)	0.90 ( 44.09)	20.84 (0.62)	0.25 (0.18)	0.09 (0.12)	1.80	979.53
Triangulo	15	0.5	2.34 (0.30)	1.04 ( 69.80)	21.07 (0.58)	0.25 (0.19)	0.10 (0.13)	1.83	967.41
Triangulo	15	2.0	3.06 (0.48)	1.90 (273.89)	28.38 (2.56)	0.46 (0.34)	0.33 (0.45)	1.66	967.31
Triangulo	15	5.0	2.41 (0.34)	0.89 ( 48.27)	21.65 (0.58)	0.22 (0.15)	0.07 (0.10)	1.82	976.09
Triangulo	30	0.2	2.33 (0.33)	0.88 ( 42.94)	21.08 (0.62)	0.23 (0.17)	0.08 (0.10)	1.84	974.80
Triangulo	30	0.5	2.37 (0.32)	1.16 ( 84.70)	21.52 (0.59)	0.30 (0.22)	0.14 (0.19)	1.77	971.19
Triangulo	30	2.0	3.53 (0.54)	2.42 (269.64)	33.81 (3.07)	0.54 (0.42)	0.46 (0.62)	1.65	973.61
Triangulo	30	5.0	2.28 (0.32)	0.92 ( 46.95)	20.67 (0.59)	0.22 (0.16)	0.07 (0.10)	1.82	969.20
Cuadrado	0	1.0	2.89 (0.37)	1.04 ( 57.16)	26.69 (0.74)	0.25 (0.18)	0.09 (0.14)	1.83	1196.02
Cuadrado	5	0.2	2.84 (0.39)	1.04 ( 54.11)	26.64 (0.74)	0.26 (0.18)	0.10 (0.13)	1.83	1067.11
Cuadrado	5	0.5	2.89 (0.39)	1.12 ( 78.45)	26.78 (0.75)	0.29 (0.22)	0.14 (0.18)	1.77	1052.42
Cuadrado	5	2.0	3.23 (0.50)	1.71 (361.07)	30.57 (2.78)	0.47 (0.33)	0.33 (0.46)	1.65	1062.69
Cuadrado	5	5.0	2.93 (0.36)	1.03 ( 61.57)	27.17 (0.69)	0.28 (0.19)	0.11 (0.14)	1.78	1067.61
Cuadrado	15	0.2	2.87 (0.33)	1.02 ( 50.94)	26.85 (0.74)	0.26 (0.21)	0.12 (0.18)	1.81	1083.59
Cuadrado	15	0.5	2.90 (0.37)	1.26 (113.64)	26.95 (0.84)	0.33 (0.24)	0.17 (0.21)	1.76	1085.62
Cuadrado	15	2.0	3.82 (0.62)	2.47 (352.86)	36.83 (3.31)	0.56 (0.42)	0.49 (0.65)	1.68	1084.53
Cuadrado	15	5.0	2.97 (0.38)	0.99 ( 53.02)	27.83 (0.71)	0.25 (0.18)	0.09 (0.12)	1.82	1058.03
Cuadrado	30	0.2	2.85 (0.41)	0.99 ( 54.17)	27.13 (0.66)	0.25 (0.19)	0.10 (0.13)	1.81	1054.78
Cuadrado	30	0.5	2.90 (0.37)	1.41 (126.36)	27.36 (0.79)	0.39 (0.27)	0.22 (0.27)	1.74	1060.97
Cuadrado	30	2.0	4.42 (0.66)	3.09 (359.04)	43.17 (3.79)	0.75 (0.52)	0.83 (1.01)	1.52	1068.39
Cuadrado	30	5.0	2.82 (0.36)	1.04 ( 51.67)	26.62 (0.78)	0.29 (0.19)	0.12 (0.15)	1.80	963.23
Ajedrez	0	1.0	3.90 (0.57)	1.35 ( 86.22)	37.70 (1.14)	0.40 (0.28)	0.24 (0.30)	1.69	1159.66
Ajedrez	5	0.2	3.88 (0.50)	1.35 ( 74.61)	37.43 (1.21)	0.37 (0.28)	0.21 (0.29)	1.75	1145.73
Ajedrez	5	0.5	3.87 (0.51)	1.48 (137.72)	37.56 (1.03)	0.42 (0.30)	0.27 (0.33)	1.71	1085.86
Ajedrez	5	2.0	4.32 (0.69)	2.31 (496.40)	42.10 (3.58)	0.53 (0.41)	0.45 (0.63)	1.64	1094.86
Ajedrez	5	5.0	3.95 (0.48)	1.33 ( 86.14)	38.28 (1.01)	0.37 (0.25)	0.20 (0.23)	1.73	1088.11
Ajedrez	15	0.2	3.85 (0.51)	1.32 ( 87.40)	37.65 (1.08)	0.35 (0.24)	0.18 (0.22)	1.73	1048.25
Ajedrez	15	0.5	3.91 (0.50)	1.68 (148.48)	37.68 (1.06)	0.45 (0.31)	0.29 (0.39)	1.73	1041.73
Ajedrez	15	2.0	5.16 (0.79)	3.33 (470.46)	50.62 (4.20)	0.70 (0.46)	0.70 (0.84)	1.51	1052.42
Ajedrez	15	5.0	4.05 (0.51)	1.27 ( 89.61)	39.15 (1.05)	0.32 (0.25)	0.17 (0.23)	1.77	1017.16
Ajedrez	30	0.2	3.93 (0.54)	1.27 ( 82.05)	38.18 (0.98)	0.34 (0.23)	0.17 (0.21)	1.78	953.34
Ajedrez	30	0.5	3.96 (0.49)	1.91 (153.05)	38.22 (1.03)	0.46 (0.33)	0.32 (0.41)	1.75	960.00
Ajedrez	30	2.0	6.16 (0.93)	4.24 (502.15)	59.80 (5.41)	0.81 (0.61)	1.02 (1.48)	1.66	966.78
Ajedrez	30	5.0	3.87 (0.51)	1.37 ( 85.73)	37.34 (1.08)	0.36 (0.27)	0.20 (0.25)	1.74	977.03

**Tabla A-1.:** Promedio y desviación estándar (en paréntesis) de  $RMPSE$ ,  $d(\beta, \hat{\beta})$ ,  $d_2(\beta_v, \hat{\beta}_v)$ ,  $d(\theta, \hat{\theta})$ ,  $d_2(\theta, \hat{\theta})$ ,  $\hat{\theta}$  y tiempo total de simulación de 200 replicaciones. Modelo de respuesta continua,  $n = 100$ , bajo niveles de contaminación  $\gamma_c$  y factor de contaminación de magnitud  $\gamma_b$ .

$\beta_0$	$\gamma_c$ (%)	$\gamma_b$	<i>RMSPE</i>	$d(\beta, \hat{\beta})$	$d_2(\beta_v, \hat{\beta}_v)$	$d(\theta, \hat{\theta})$	$d_2(\theta, \hat{\theta})$	$\hat{\theta}$	Tiempo (Seg)
T	0	1.0	2.93 (0.29)	1.26 ( 47.00)	27.84 (0.76)	0.21 (0.15)	0.07 (0.09)	1.86	1783.34
T	5	0.2	2.96 (0.28)	1.25 ( 53.26)	27.63 (0.77)	0.21 (0.16)	0.07 (0.10)	1.86	1498.73
T	5	0.5	2.97 (0.30)	1.34 ( 76.70)	27.78 (0.78)	0.21 (0.17)	0.08 (0.11)	1.85	1506.91
T	5	2.0	3.43 (0.41)	2.01 (354.59)	32.35 (2.90)	0.34 (0.27)	0.19 (0.30)	1.81	1518.92
T	5	5.0	3.01 (0.31)	1.25 ( 49.62)	28.49 (0.76)	0.20 (0.14)	0.06 (0.08)	1.87	1537.73
T	15	0.2	2.96 (0.29)	1.23 ( 45.64)	27.93 (0.77)	0.20 (0.15)	0.06 (0.08)	1.88	1561.23
T	15	0.5	3.00 (0.30)	1.51 ( 95.55)	28.18 (0.88)	0.28 (0.20)	0.12 (0.14)	1.82	1590.94
T	15	2.0	4.33 (0.55)	3.05 (362.01)	42.39 (3.78)	0.44 (0.34)	0.31 (0.43)	1.76	1652.23
T	15	5.0	3.10 (0.33)	1.22 ( 51.14)	29.48 (0.73)	0.20 (0.15)	0.06 (0.08)	1.89	1668.61
T	30	0.2	2.99 (0.29)	1.19 ( 46.70)	28.36 (0.76)	0.19 (0.16)	0.06 (0.09)	1.87	1649.62
T	30	0.5	3.04 (0.33)	1.71 (103.79)	28.77 (0.77)	0.28 (0.20)	0.12 (0.16)	1.85	1654.70
T	30	2.0	5.21 (0.71)	3.84 (341.59)	51.19 (4.05)	0.57 (0.44)	0.52 (0.67)	1.80	1672.66
T	30	5.0	2.92 (0.28)	1.25 ( 50.57)	27.42 (0.82)	0.18 (0.14)	0.05 (0.07)	1.89	1554.34
Triangulo	0	1.0	2.13 (0.20)	0.96 ( 45.42)	18.86 (0.58)	0.15 (0.11)	0.04 (0.05)	1.89	2003.67
Triangulo	5	0.2	2.10 (0.22)	0.96 ( 42.75)	18.72 (0.58)	0.16 (0.12)	0.04 (0.06)	1.89	1701.25
Triangulo	5	0.5	2.14 (0.22)	1.01 ( 45.46)	18.90 (0.64)	0.18 (0.12)	0.05 (0.05)	1.86	1553.44
Triangulo	5	2.0	2.49 (0.33)	1.47 (213.83)	22.89 (2.01)	0.25 (0.19)	0.09 (0.13)	1.84	1636.06
Triangulo	5	5.0	2.19 (0.23)	0.96 ( 42.75)	19.54 (0.70)	0.15 (0.11)	0.03 (0.04)	1.90	1670.47
Triangulo	15	0.2	2.16 (0.22)	0.95 ( 43.29)	18.99 (0.62)	0.17 (0.13)	0.04 (0.06)	1.89	1675.58
Triangulo	15	0.5	2.16 (0.22)	1.11 ( 60.90)	19.35 (0.62)	0.16 (0.12)	0.04 (0.05)	1.92	1655.02
Triangulo	15	2.0	3.10 (0.41)	2.13 (263.28)	29.47 (2.83)	0.35 (0.27)	0.20 (0.29)	1.76	1656.39
Triangulo	15	5.0	2.27 (0.23)	0.96 ( 45.49)	20.42 (0.57)	0.17 (0.11)	0.04 (0.05)	1.90	1640.53
Triangulo	30	0.2	2.18 (0.22)	0.94 ( 43.99)	19.46 (0.62)	0.15 (0.10)	0.03 (0.04)	1.89	1646.22
Triangulo	30	0.5	2.25 (0.26)	1.23 ( 70.19)	19.91 (0.61)	0.22 (0.15)	0.07 (0.09)	1.88	1627.42
Triangulo	30	2.0	3.71 (0.50)	2.69 (221.26)	35.72 (2.70)	0.43 (0.33)	0.29 (0.40)	1.76	1625.55
Triangulo	30	5.0	2.09 (0.23)	0.97 ( 45.11)	18.66 (0.65)	0.15 (0.11)	0.03 (0.04)	1.89	1512.52
Cuadrado	0	1.0	2.61 (0.28)	1.11 ( 43.52)	24.16 (0.79)	0.18 (0.13)	0.05 (0.06)	1.88	1951.70
Cuadrado	5	0.2	2.63 (0.29)	1.10 ( 40.22)	24.07 (0.72)	0.18 (0.12)	0.04 (0.06)	1.89	1780.00
Cuadrado	5	0.5	2.61 (0.29)	1.19 ( 68.86)	24.14 (0.77)	0.21 (0.16)	0.07 (0.10)	1.88	1756.22
Cuadrado	5	2.0	3.05 (0.37)	1.82 (288.77)	28.70 (2.49)	0.31 (0.22)	0.14 (0.18)	1.85	1734.34
Cuadrado	5	5.0	2.66 (0.26)	1.10 ( 47.30)	24.91 (0.72)	0.18 (0.14)	0.05 (0.08)	1.90	1726.38
Cuadrado	15	0.2	2.61 (0.26)	1.08 ( 45.04)	24.34 (0.74)	0.17 (0.13)	0.05 (0.06)	1.88	1941.64
Cuadrado	15	0.5	2.68 (0.26)	1.32 ( 82.49)	24.58 (0.68)	0.23 (0.17)	0.08 (0.10)	1.83	1638.17
Cuadrado	15	2.0	3.84 (0.49)	2.67 (347.98)	36.97 (3.62)	0.44 (0.33)	0.30 (0.44)	1.73	1546.56
Cuadrado	15	5.0	2.77 (0.28)	1.08 ( 48.63)	25.91 (0.67)	0.17 (0.13)	0.04 (0.06)	1.89	1541.91
Cuadrado	30	0.2	2.66 (0.27)	1.06 ( 44.80)	24.80 (0.66)	0.17 (0.13)	0.05 (0.06)	1.88	1637.09
Cuadrado	30	0.5	2.72 (0.27)	1.49 ( 83.98)	25.12 (0.74)	0.26 (0.18)	0.10 (0.12)	1.85	1636.52
Cuadrado	30	2.0	4.64 (0.53)	3.40 (261.09)	45.34 (3.25)	0.55 (0.39)	0.45 (0.57)	1.77	1694.30
Cuadrado	30	5.0	2.57 (0.27)	1.10 ( 41.59)	23.86 (0.79)	0.20 (0.14)	0.06 (0.08)	1.85	1692.61
Ajedrez	0	1.0	3.64 (0.38)	1.43 ( 64.78)	34.60 (1.07)	0.25 (0.20)	0.10 (0.14)	1.82	1744.61
Ajedrez	5	0.2	3.54 (0.37)	1.42 ( 61.15)	34.28 (1.07)	0.23 (0.17)	0.08 (0.12)	1.86	1595.39
Ajedrez	5	0.5	3.54 (0.38)	1.56 (100.25)	33.96 (1.16)	0.27 (0.18)	0.11 (0.12)	1.84	1606.56
Ajedrez	5	2.0	4.18 (0.54)	2.49 (445.01)	40.03 (3.53)	0.35 (0.26)	0.19 (0.26)	1.83	1640.75
Ajedrez	5	5.0	3.66 (0.37)	1.41 ( 63.56)	35.36 (1.03)	0.24 (0.15)	0.08 (0.09)	1.87	1631.48
Ajedrez	15	0.2	3.61 (0.34)	1.38 ( 60.20)	34.73 (1.06)	0.24 (0.18)	0.09 (0.12)	1.86	1691.77
Ajedrez	15	0.5	3.62 (0.36)	1.77 (126.42)	34.40 (1.10)	0.29 (0.23)	0.14 (0.20)	1.83	1695.98
Ajedrez	15	2.0	5.25 (0.73)	3.75 (468.32)	52.17 (4.85)	0.52 (0.38)	0.41 (0.55)	1.81	1590.23
Ajedrez	15	5.0	3.84 (0.36)	1.37 ( 69.10)	36.81 (1.02)	0.23 (0.18)	0.09 (0.12)	1.85	1526.25
Ajedrez	30	0.2	3.67 (0.39)	1.35 ( 63.88)	35.15 (1.09)	0.23 (0.17)	0.08 (0.10)	1.86	1516.06
Ajedrez	30	0.5	3.61 (0.38)	2.03 (121.48)	34.99 (1.03)	0.33 (0.25)	0.17 (0.23)	1.81	1513.16
Ajedrez	30	2.0	6.47 (0.81)	4.76 (430.62)	63.68 (5.35)	0.68 (0.53)	0.74 (1.01)	1.84	1518.03
Ajedrez	30	5.0	3.53 (0.39)	1.43 ( 59.90)	34.20 (1.10)	0.26 (0.19)	0.10 (0.14)	1.84	1609.34

**Tabla A-2.:** Promedio y desviación estándar (en paréntesis) de *RMPSE*,  $d(\beta, \hat{\beta})$ ,  $d_2(\beta_v, \hat{\beta}_v)$ ,  $d(\theta, \hat{\theta})$ ,  $d_2(\theta, \hat{\theta})$ ,  $\hat{\theta}$  y tiempo total de simulación de 200 repeticiones. Modelo de respuesta continua,  $n = 170$ , bajo niveles de contaminación  $\gamma_c$  y factor de contaminación de magnitud  $\gamma_b$ .

$\beta_0$	$\gamma_c$ (%)	$\gamma_b$	$RMSPE$	$d(\beta, \hat{\beta})$	$d_2(\beta_v, \hat{\beta}_v)$	$d(\theta, \hat{\theta})$	$d_2(\theta, \hat{\theta})$	$\hat{\theta}$	tiempo
T	0	1.0	2.77 (0.27)	1.27 ( 45.63)	25.77 (0.77)	0.17 (0.13)	0.04 (0.06)	1.90	1638.36
T	5	0.2	2.80 (0.23)	1.28 ( 46.02)	26.25 (0.69)	0.17 (0.13)	0.05 (0.07)	1.91	1566.25
T	5	0.5	2.78 (0.25)	1.28 ( 46.61)	26.06 (0.68)	0.17 (0.12)	0.04 (0.06)	1.89	1492.77
T	5	2.0	2.83 (0.25)	1.38 ( 64.02)	26.17 (0.77)	0.20 (0.13)	0.06 (0.06)	1.90	1702.75
T	5	5.0	3.46 (0.43)	2.22 (341.10)	32.77 (3.18)	0.27 (0.20)	0.11 (0.15)	1.88	1747.69
T	15	0.2	2.88 (0.26)	1.28 ( 47.67)	27.17 (0.70)	0.17 (0.12)	0.04 (0.06)	1.89	1742.16
T	15	0.5	2.82 (0.25)	1.26 ( 47.38)	26.35 (0.79)	0.17 (0.14)	0.05 (0.07)	1.89	1728.25
T	15	2.0	2.86 (0.28)	1.56 ( 87.90)	26.65 (0.75)	0.21 (0.17)	0.07 (0.11)	1.87	1722.27
T	15	5.0	4.49 (0.55)	3.24 (312.63)	43.46 (3.53)	0.40 (0.30)	0.25 (0.33)	1.77	1728.06
T	30	0.2	3.01 (0.30)	1.27 ( 53.09)	28.43 (0.75)	0.18 (0.13)	0.05 (0.06)	1.89	1727.69
T	30	0.5	2.92 (0.28)	1.23 ( 49.47)	26.99 (0.76)	0.18 (0.13)	0.05 (0.06)	1.88	1768.66
T	30	2.0	2.92 (0.24)	1.75 (101.72)	27.54 (0.82)	0.23 (0.17)	0.08 (0.11)	1.88	1791.23
T	30	5.0	5.42 (0.58)	4.06 (310.35)	52.90 (3.87)	0.54 (0.38)	0.44 (0.53)	1.77	1809.94
Triangulo	0	1.0	2.00 (0.19)	0.98 ( 39.24)	17.46 (0.60)	0.14 (0.10)	0.03 (0.04)	1.91	1758.42
Triangulo	5	0.2	2.03 (0.20)	0.98 ( 38.32)	17.84 (0.62)	0.13 (0.10)	0.03 (0.04)	1.92	1790.72
Triangulo	5	0.5	2.03 (0.19)	0.97 ( 36.62)	17.63 (0.67)	0.13 (0.09)	0.03 (0.03)	1.93	1601.69
Triangulo	5	2.0	2.03 (0.18)	1.04 ( 47.08)	17.86 (0.60)	0.14 (0.11)	0.03 (0.04)	1.93	1611.00
Triangulo	5	5.0	2.50 (0.29)	1.59 (233.46)	22.98 (2.34)	0.20 (0.15)	0.07 (0.09)	1.87	1622.97
Triangulo	15	0.2	2.11 (0.20)	0.99 ( 44.37)	18.59 (0.65)	0.13 (0.09)	0.02 (0.03)	1.92	1627.64
Triangulo	15	0.5	2.05 (0.21)	0.97 ( 40.97)	17.97 (0.64)	0.14 (0.10)	0.03 (0.04)	1.92	1624.75
Triangulo	15	2.0	2.12 (0.19)	1.14 ( 53.95)	18.49 (0.57)	0.15 (0.11)	0.03 (0.05)	1.92	1612.56
Triangulo	15	5.0	3.17 (0.37)	2.26 (214.39)	30.16 (2.50)	0.27 (0.22)	0.12 (0.20)	1.87	1649.80
Triangulo	30	0.2	2.20 (0.20)	1.01 ( 41.15)	19.70 (0.66)	0.11 (0.08)	0.02 (0.02)	1.93	1740.44
Triangulo	30	0.5	2.11 (0.21)	0.96 ( 38.74)	18.54 (0.58)	0.13 (0.09)	0.02 (0.04)	1.93	1774.62
Triangulo	30	2.0	2.15 (0.19)	1.26 ( 63.45)	19.12 (0.61)	0.17 (0.11)	0.04 (0.05)	1.89	1768.73
Triangulo	30	5.0	3.86 (0.48)	2.80 (229.72)	36.75 (2.92)	0.42 (0.31)	0.27 (0.34)	1.77	1768.55
Cuadrado	0	1.0	2.44 (0.22)	1.12 ( 43.06)	22.21 (0.77)	0.15 (0.11)	0.03 (0.04)	1.90	1742.56
Cuadrado	5	0.2	2.46 (0.22)	1.13 ( 46.24)	22.66 (0.77)	0.16 (0.11)	0.04 (0.05)	1.90	1894.61
Cuadrado	5	0.5	2.46 (0.24)	1.12 ( 44.62)	22.49 (0.63)	0.14 (0.11)	0.03 (0.05)	1.92	1767.09
Cuadrado	5	2.0	2.50 (0.23)	1.21 ( 59.39)	22.70 (0.74)	0.16 (0.13)	0.04 (0.06)	1.90	1786.00
Cuadrado	5	5.0	3.07 (0.34)	1.96 (296.06)	28.76 (2.71)	0.27 (0.18)	0.10 (0.13)	1.85	1797.91
Cuadrado	15	0.2	2.56 (0.24)	1.14 ( 45.38)	23.67 (0.82)	0.15 (0.10)	0.03 (0.04)	1.91	1795.80
Cuadrado	15	0.5	2.49 (0.21)	1.11 ( 44.33)	22.89 (0.73)	0.15 (0.12)	0.03 (0.05)	1.89	1784.03
Cuadrado	15	2.0	2.52 (0.24)	1.36 ( 83.58)	23.14 (0.76)	0.19 (0.14)	0.06 (0.08)	1.89	1779.19
Cuadrado	15	5.0	3.96 (0.44)	2.81 (292.53)	37.88 (3.16)	0.37 (0.29)	0.22 (0.31)	1.80	1745.72
Cuadrado	30	0.2	2.69 (0.25)	1.13 ( 45.97)	24.92 (0.73)	0.16 (0.11)	0.04 (0.05)	1.88	1660.36
Cuadrado	30	0.5	2.53 (0.23)	1.09 ( 44.59)	23.41 (0.73)	0.14 (0.10)	0.03 (0.04)	1.92	1750.12
Cuadrado	30	2.0	2.62 (0.24)	1.54 ( 76.43)	24.03 (0.74)	0.21 (0.14)	0.06 (0.08)	1.89	1754.88
Cuadrado	30	5.0	4.79 (0.56)	3.56 (269.84)	46.75 (3.62)	0.52 (0.35)	0.40 (0.48)	1.76	1764.27
Ajedrez	0	1.0	3.33 (0.33)	1.45 ( 48.57)	32.08 (1.14)	0.20 (0.15)	0.06 (0.08)	1.89	1262.19
Ajedrez	5	0.2	3.42 (0.29)	1.46 ( 54.71)	32.61 (1.09)	0.20 (0.16)	0.07 (0.11)	1.90	1457.72
Ajedrez	5	0.5	3.38 (0.30)	1.44 ( 54.80)	32.45 (1.20)	0.20 (0.15)	0.06 (0.08)	1.87	1414.86
Ajedrez	5	2.0	3.33 (0.32)	1.60 ( 88.90)	31.88 (1.08)	0.22 (0.16)	0.07 (0.09)	1.89	1199.75
Ajedrez	5	5.0	4.13 (0.53)	2.68 (450.74)	40.07 (4.14)	0.31 (0.23)	0.15 (0.19)	1.90	1206.03
Ajedrez	15	0.2	3.51 (0.30)	1.45 ( 52.19)	33.95 (1.05)	0.22 (0.15)	0.07 (0.09)	1.88	1249.84
Ajedrez	15	0.5	3.46 (0.31)	1.42 ( 52.40)	32.73 (1.02)	0.21 (0.14)	0.06 (0.08)	1.88	1268.86
Ajedrez	15	2.0	3.40 (0.31)	1.83 (111.83)	32.40 (1.14)	0.24 (0.18)	0.09 (0.13)	1.89	1267.16
Ajedrez	15	5.0	5.42 (0.67)	3.97 (438.10)	53.27 (4.86)	0.47 (0.34)	0.33 (0.45)	1.85	1277.28
Ajedrez	30	0.2	3.70 (0.35)	1.43 ( 66.00)	35.41 (1.03)	0.22 (0.14)	0.07 (0.08)	1.87	1270.95
Ajedrez	30	0.5	3.52 (0.34)	1.37 ( 55.03)	33.50 (1.01)	0.19 (0.14)	0.06 (0.08)	1.90	1275.75
Ajedrez	30	2.0	3.51 (0.30)	2.09 (104.33)	33.40 (0.99)	0.27 (0.20)	0.11 (0.16)	1.89	1270.16
Ajedrez	30	5.0	6.69 (0.78)	5.01 (402.28)	65.67 (4.88)	0.55 (0.42)	0.48 (0.70)	1.96	1273.97

**Tabla A-3.:** Promedio y desviación estándar (en paréntesis) de  $RMPSE$ ,  $d(\beta, \hat{\beta})$ ,  $d_2(\beta_v, \hat{\beta}_v)$ ,  $d(\theta, \hat{\theta})$ ,  $d_2(\theta, \hat{\theta})$ ,  $\hat{\theta}$  y tiempo total de simulación de 200 replicaciones. Modelo de respuesta continua,  $n = 220$ , bajo niveles de contaminación  $\gamma_c$  y factor de contaminación de magnitud  $\gamma_b$ .

$\beta_0$	$\gamma_c$ (%)	$\gamma_b$	$ABC$	$d(\beta, \hat{\beta})$	$d_2(\beta_v, \hat{\beta}_v)$	$d(\theta, \hat{\theta})$	$d_2(\theta, \hat{\theta})$	$\hat{\theta}$	Tiempo (Seg)
T	0	1.0	0.64 (0.10)	1.34 (62.05)	33.66 (0.75)	1.02 (0.34)	1.16 (0.70)	0.98	846.77
T	5	0.2	0.67 (0.12)	1.34 (60.55)	33.65 (0.71)	1.07 (0.38)	1.30 (0.85)	0.93	811.22
T	5	0.5	0.65 (0.10)	1.33 (56.28)	33.51 (0.71)	1.06 (0.37)	1.26 (0.77)	0.94	1094.06
T	5	2.0	0.65 (0.11)	1.33 (57.10)	33.45 (0.77)	1.13 (0.37)	1.41 (0.82)	0.87	1189.36
T	5	5.0	0.64 (0.11)	1.34 (61.30)	34.07 (0.77)	0.84 (0.37)	0.83 (0.63)	1.17	1357.03
T	15	0.2	0.65 (0.12)	1.34 (66.09)	33.72 (0.80)	0.93 (0.36)	1.00 (0.72)	1.07	1404.56
T	15	0.5	0.66 (0.12)	1.34 (63.88)	33.45 (0.79)	1.13 (0.35)	1.40 (0.82)	0.87	1412.08
T	15	2.0	0.65 (0.12)	1.34 (59.69)	33.42 (0.81)	1.20 (0.36)	1.57 (0.84)	0.80	1396.25
T	15	5.0	0.62 (0.12)	1.34 (65.36)	34.68 (0.83)	0.70 (0.37)	0.63 (0.57)	1.30	1407.11
T	30	0.2	0.62 (0.12)	1.33 (70.95)	34.01 (0.80)	0.82 (0.36)	0.79 (0.62)	1.19	1404.72
T	30	0.5	0.65 (0.12)	1.34 (59.30)	33.34 (0.76)	1.28 (0.33)	1.75 (0.84)	0.72	1399.09
T	30	2.0	0.66 (0.11)	1.36 (54.54)	33.36 (0.72)	1.43 (0.34)	2.16 (0.96)	0.57	1377.50
T	30	5.0	0.65 (0.12)	1.34 (61.84)	33.58 (0.76)	1.07 (0.38)	1.29 (0.78)	0.93	1404.30
Triangulo	0	1.0	0.61 (0.13)	1.31 (64.46)	25.98 (0.69)	0.60 (0.34)	0.48 (0.48)	1.41	1478.14
Triangulo	5	0.2	0.60 (0.12)	1.30 (69.26)	25.91 (0.65)	0.61 (0.35)	0.49 (0.50)	1.41	1550.36
Triangulo	5	0.5	0.60 (0.12)	1.31 (63.19)	25.78 (0.67)	0.66 (0.37)	0.57 (0.53)	1.35	1413.61
Triangulo	5	2.0	0.62 (0.14)	1.31 (63.27)	25.82 (0.71)	0.66 (0.34)	0.55 (0.52)	1.35	1416.91
Triangulo	5	5.0	0.58 (0.13)	1.30 (76.14)	26.15 (0.70)	0.53 (0.33)	0.39 (0.43)	1.50	1406.84
Triangulo	15	0.2	0.59 (0.13)	1.30 (68.14)	26.01 (0.66)	0.52 (0.34)	0.38 (0.43)	1.51	1417.55
Triangulo	15	0.5	0.60 (0.14)	1.31 (63.83)	25.77 (0.63)	0.76 (0.36)	0.71 (0.59)	1.24	1416.98
Triangulo	15	2.0	0.60 (0.12)	1.33 (57.22)	25.68 (0.69)	0.86 (0.34)	0.86 (0.62)	1.14	1173.89
Triangulo	15	5.0	0.59 (0.13)	1.28 (72.80)	26.37 (0.71)	0.42 (0.27)	0.26 (0.29)	1.65	786.17
Triangulo	30	0.2	0.58 (0.13)	1.29 (74.61)	26.12 (0.74)	0.48 (0.33)	0.34 (0.38)	1.55	957.75
Triangulo	30	0.5	0.60 (0.12)	1.32 (62.24)	25.66 (0.63)	0.86 (0.35)	0.86 (0.62)	1.14	955.62
Triangulo	30	2.0	0.60 (0.13)	1.34 (61.07)	25.66 (0.64)	1.06 (0.36)	1.25 (0.77)	0.94	941.52
Triangulo	30	5.0	0.60 (0.14)	1.30 (66.26)	25.82 (0.66)	0.60 (0.36)	0.48 (0.50)	1.41	1044.81
Cuadrado	0	1.0	0.64 (0.11)	1.28 (65.90)	30.38 (0.81)	0.88 (0.38)	0.91 (0.71)	1.13	1178.97
Cuadrado	5	0.2	0.65 (0.11)	1.28 (63.09)	30.39 (0.80)	0.92 (0.40)	1.00 (0.79)	1.08	1072.59
Cuadrado	5	0.5	0.64 (0.12)	1.28 (57.11)	30.19 (0.79)	1.04 (0.36)	1.20 (0.75)	0.97	885.75
Cuadrado	5	2.0	0.66 (0.12)	1.28 (55.36)	30.12 (0.77)	1.02 (0.34)	1.16 (0.71)	0.98	1022.94
Cuadrado	5	5.0	0.63 (0.12)	1.28 (61.19)	30.87 (0.75)	0.75 (0.35)	0.68 (0.53)	1.26	1003.97
Cuadrado	15	0.2	0.66 (0.11)	1.27 (57.03)	30.46 (0.80)	0.82 (0.34)	0.79 (0.60)	1.18	1048.83
Cuadrado	15	0.5	0.67 (0.12)	1.29 (61.57)	30.12 (0.78)	1.05 (0.38)	1.24 (0.82)	0.95	957.66
Cuadrado	15	2.0	0.66 (0.12)	1.30 (59.58)	30.15 (0.74)	1.12 (0.36)	1.39 (0.79)	0.88	951.73
Cuadrado	15	5.0	0.61 (0.12)	1.27 (64.52)	31.28 (0.82)	0.61 (0.35)	0.50 (0.51)	1.40	1007.80
Cuadrado	30	0.2	0.63 (0.12)	1.27 (67.78)	30.77 (0.89)	0.69 (0.36)	0.61 (0.55)	1.32	919.92
Cuadrado	30	0.5	0.65 (0.12)	1.29 (62.84)	29.99 (0.78)	1.13 (0.37)	1.41 (0.86)	0.87	900.77
Cuadrado	30	2.0	0.66 (0.12)	1.30 (54.79)	30.09 (0.78)	1.29 (0.36)	1.78 (0.92)	0.71	989.69
Cuadrado	30	5.0	0.66 (0.12)	1.28 (63.39)	30.17 (0.80)	0.95 (0.36)	1.03 (0.71)	1.05	1013.03
Ajedrez	0	1.0	0.67 (0.11)	1.30 (56.05)	41.58 (0.77)	1.20 (0.31)	1.52 (0.71)	0.80	1029.00
Ajedrez	5	0.2	0.67 (0.11)	1.29 (55.87)	41.38 (0.78)	1.27 (0.35)	1.73 (0.89)	0.73	1086.31
Ajedrez	5	0.5	0.66 (0.11)	1.29 (56.77)	41.34 (0.74)	1.30 (0.33)	1.80 (0.84)	0.70	1123.33
Ajedrez	5	2.0	0.66 (0.11)	1.29 (58.42)	41.38 (0.76)	1.32 (0.33)	1.86 (0.89)	0.68	1323.02
Ajedrez	5	5.0	0.66 (0.11)	1.30 (54.11)	42.03 (0.81)	1.15 (0.31)	1.41 (0.73)	0.85	1374.88
Ajedrez	15	0.2	0.67 (0.11)	1.29 (52.36)	41.51 (0.78)	1.18 (0.34)	1.51 (0.80)	0.82	1352.72
Ajedrez	15	0.5	0.67 (0.11)	1.29 (60.36)	41.19 (0.77)	1.39 (0.31)	2.03 (0.85)	0.61	1326.36
Ajedrez	15	2.0	0.67 (0.11)	1.30 (55.49)	41.31 (0.81)	1.41 (0.32)	2.10 (0.90)	0.59	1249.31
Ajedrez	15	5.0	0.64 (0.12)	1.29 (62.04)	42.49 (0.77)	0.86 (0.33)	0.85 (0.60)	1.14	1213.45
Ajedrez	30	0.2	0.67 (0.12)	1.29 (60.60)	41.78 (0.77)	1.07 (0.35)	1.28 (0.78)	0.93	1289.50
Ajedrez	30	0.5	0.67 (0.12)	1.30 (53.92)	41.28 (0.80)	1.41 (0.32)	2.09 (0.93)	0.59	1287.50
Ajedrez	30	2.0	0.68 (0.12)	1.31 (52.62)	41.19 (0.72)	1.55 (0.32)	2.52 (0.99)	0.45	1285.89
Ajedrez	30	5.0	0.67 (0.11)	1.30 (57.32)	41.34 (0.69)	1.33 (0.32)	1.86 (0.87)	0.67	1257.70

**Tabla A-4.:** Promedio y desviación estándar (en paréntesis) de  $ABC$ ,  $d(\beta, \hat{\beta})$ ,  $d_2(\beta_v, \hat{\beta}_v)$ ,  $d(\theta, \hat{\theta})$ ,  $d_2(\theta, \hat{\theta})$ ,  $\hat{\theta}$  y tiempo total de simulación de 200 replicaciones. Modelo de respuesta binaria,  $n = 100$ , bajo niveles de contaminación  $\gamma_c$  y factor de contaminación de magnitud  $\gamma_b$ .

$\beta_0$	$\gamma_c$ (%)	$\gamma_b$	$ABC$	$d(\beta, \hat{\beta})$	$d_2(\beta_v, \hat{\beta}_v)$	$d(\theta, \hat{\theta})$	$d_2(\theta, \hat{\theta})$	$\hat{\theta}$	Tiempo (Seg)	
T		0	1.0	0.67 (0.08)	1.54 (140.87)	33.00 (0.93)	0.81 (0.31)	0.75 (0.51)	1.19	1926.41
T		5	0.2	0.68 (0.08)	1.52 (136.68)	32.81 (0.76)	0.84 (0.31)	0.79 (0.51)	1.17	2125.33
T		5	0.5	0.68 (0.08)	1.54 (144.42)	32.67 (0.80)	0.88 (0.30)	0.86 (0.54)	1.12	2187.80
T		5	2.0	0.68 (0.09)	1.54 (143.92)	32.70 (0.82)	0.89 (0.30)	0.88 (0.56)	1.11	2134.94
T		5	5.0	0.66 (0.09)	1.52 (131.46)	33.44 (0.92)	0.63 (0.29)	0.48 (0.40)	1.39	2071.56
T		15	0.2	0.67 (0.08)	1.52 (137.14)	33.02 (0.87)	0.75 (0.29)	0.65 (0.44)	1.25	2097.56
T		15	0.5	0.68 (0.08)	1.53 (141.51)	32.59 (0.84)	0.93 (0.32)	0.96 (0.58)	1.08	2143.17
T		15	2.0	0.68 (0.08)	1.57 (158.42)	32.73 (0.86)	0.97 (0.30)	1.03 (0.57)	1.03	2218.36
T		15	5.0	0.66 (0.09)	1.50 (120.13)	34.19 (0.95)	0.47 (0.26)	0.29 (0.27)	1.54	2146.94
T		30	0.2	0.67 (0.08)	1.53 (144.45)	33.43 (0.90)	0.64 (0.27)	0.49 (0.38)	1.36	2073.75
T		30	0.5	0.69 (0.09)	1.56 (156.94)	32.49 (0.99)	1.05 (0.33)	1.21 (0.66)	0.95	2039.44
T		30	2.0	0.70 (0.08)	1.58 (154.70)	32.58 (0.89)	1.15 (0.28)	1.40 (0.65)	0.85	2032.20
T		30	5.0	0.68 (0.08)	1.53 (148.45)	32.73 (0.92)	0.85 (0.29)	0.80 (0.49)	1.15	2063.20
Triangulo		0	1.0	0.63 (0.10)	1.47 (113.17)	25.97 (0.91)	0.46 (0.25)	0.28 (0.25)	1.56	2367.30
Triangulo		5	0.2	0.64 (0.10)	1.45 (107.98)	25.78 (0.95)	0.43 (0.23)	0.24 (0.22)	1.59	2332.52
Triangulo		5	0.5	0.63 (0.09)	1.47 (113.06)	25.70 (0.85)	0.53 (0.25)	0.35 (0.30)	1.47	2289.17
Triangulo		5	2.0	0.62 (0.10)	1.48 (102.76)	25.80 (0.85)	0.56 (0.28)	0.39 (0.33)	1.44	1866.56
Triangulo		5	5.0	0.62 (0.09)	1.44 ( 83.72)	26.09 (0.76)	0.35 (0.24)	0.18 (0.22)	1.67	1907.86
Triangulo		15	0.2	0.62 (0.09)	1.44 ( 94.42)	25.91 (0.83)	0.42 (0.27)	0.25 (0.30)	1.61	1906.42
Triangulo		15	0.5	0.63 (0.09)	1.48 (113.89)	25.78 (0.94)	0.57 (0.27)	0.40 (0.33)	1.44	1694.73
Triangulo		15	2.0	0.63 (0.09)	1.49 (115.66)	25.68 (0.92)	0.66 (0.28)	0.51 (0.38)	1.34	1603.69
Triangulo		15	5.0	0.59 (0.09)	1.43 ( 76.64)	26.53 (0.76)	0.28 (0.21)	0.12 (0.16)	1.80	1518.48
Triangulo		30	0.2	0.62 (0.10)	1.44 (106.22)	26.18 (0.85)	0.32 (0.22)	0.16 (0.20)	1.72	1560.17
Triangulo		30	0.5	0.62 (0.10)	1.52 (143.31)	25.76 (1.09)	0.68 (0.27)	0.53 (0.38)	1.32	1644.98
Triangulo		30	2.0	0.64 (0.09)	1.53 (128.89)	25.74 (1.01)	0.87 (0.28)	0.83 (0.48)	1.13	1615.47
Triangulo		30	5.0	0.62 (0.09)	1.46 ( 99.49)	25.78 (0.84)	0.47 (0.27)	0.30 (0.29)	1.54	1679.94
Cuadrado		0	1.0	0.67 (0.08)	1.46 (142.37)	29.78 (0.99)	0.64 (0.29)	0.50 (0.39)	1.36	1393.97
Cuadrado		5	0.2	0.68 (0.08)	1.48 (142.65)	29.83 (0.92)	0.66 (0.29)	0.52 (0.42)	1.35	1261.00
Cuadrado		5	0.5	0.69 (0.08)	1.49 (144.30)	29.66 (0.99)	0.76 (0.29)	0.66 (0.43)	1.24	1237.42
Cuadrado		5	2.0	0.67 (0.09)	1.46 (134.84)	29.63 (0.95)	0.75 (0.30)	0.65 (0.46)	1.26	1241.42
Cuadrado		5	5.0	0.66 (0.09)	1.45 (124.55)	30.27 (0.94)	0.60 (0.27)	0.43 (0.34)	1.41	1227.61
Cuadrado		15	0.2	0.67 (0.08)	1.44 (123.06)	29.87 (0.91)	0.61 (0.28)	0.45 (0.36)	1.40	1227.23
Cuadrado		15	0.5	0.67 (0.08)	1.48 (143.11)	29.56 (0.95)	0.84 (0.29)	0.79 (0.49)	1.16	1233.34
Cuadrado		15	2.0	0.68 (0.09)	1.49 (147.46)	29.56 (0.87)	0.88 (0.30)	0.87 (0.54)	1.12	1887.47
Cuadrado		15	5.0	0.64 (0.09)	1.44 (115.94)	31.04 (0.89)	0.42 (0.23)	0.23 (0.22)	1.62	2255.19
Cuadrado		30	0.2	0.66 (0.09)	1.43 (116.25)	30.14 (0.87)	0.56 (0.27)	0.38 (0.32)	1.45	2267.41
Cuadrado		30	0.5	0.69 (0.10)	1.48 (150.24)	29.39 (1.04)	0.90 (0.31)	0.90 (0.56)	1.10	2285.17
Cuadrado		30	2.0	0.69 (0.09)	1.51 (144.56)	29.43 (0.90)	1.06 (0.30)	1.22 (0.62)	0.94	2237.22
Cuadrado		30	5.0	0.67 (0.09)	1.47 (145.44)	29.63 (0.96)	0.71 (0.28)	0.58 (0.41)	1.30	2129.23
Ajedrez		0	1.0	0.70 (0.08)	1.52 (151.60)	40.27 (0.80)	0.98 (0.28)	1.04 (0.57)	1.02	2352.33
Ajedrez		5	0.2	0.71 (0.08)	1.50 (156.54)	40.05 (0.91)	0.97 (0.27)	1.02 (0.53)	1.03	2430.38
Ajedrez		5	0.5	0.70 (0.08)	1.52 (153.77)	39.97 (0.79)	1.11 (0.32)	1.33 (0.71)	0.89	2468.22
Ajedrez		5	2.0	0.71 (0.08)	1.50 (154.61)	39.81 (0.77)	1.10 (0.28)	1.30 (0.62)	0.90	2389.00
Ajedrez		5	5.0	0.69 (0.08)	1.51 (147.37)	40.93 (0.87)	0.84 (0.27)	0.78 (0.48)	1.16	2334.22
Ajedrez		15	0.2	0.70 (0.08)	1.50 (145.83)	40.26 (0.76)	0.94 (0.29)	0.97 (0.55)	1.06	2149.48
Ajedrez		15	0.5	0.71 (0.09)	1.53 (159.16)	39.80 (0.88)	1.16 (0.28)	1.42 (0.66)	0.84	2244.34
Ajedrez		15	2.0	0.71 (0.08)	1.55 (155.86)	39.92 (0.74)	1.20 (0.29)	1.52 (0.70)	0.80	2236.05
Ajedrez		15	5.0	0.67 (0.08)	1.50 (146.61)	41.71 (0.85)	0.64 (0.29)	0.50 (0.38)	1.36	2147.38
Ajedrez		30	0.2	0.71 (0.08)	1.48 (146.92)	40.56 (0.91)	0.85 (0.28)	0.81 (0.50)	1.15	2120.53
Ajedrez		30	0.5	0.71 (0.08)	1.52 (155.12)	39.77 (0.81)	1.19 (0.32)	1.53 (0.73)	0.81	2199.86
Ajedrez		30	2.0	0.70 (0.09)	1.56 (152.07)	39.82 (0.78)	1.34 (0.29)	1.87 (0.77)	0.66	1751.69
Ajedrez		30	5.0	0.71 (0.08)	1.53 (160.32)	40.08 (0.84)	1.02 (0.31)	1.15 (0.64)	0.98	1477.91

**Tabla A-5.:** Promedio y desviación estándar (en paréntesis) de  $ABC$ ,  $d(\beta, \hat{\beta})$ ,  $d_2(\beta_v, \hat{\beta}_v)$ ,  $d(\theta, \hat{\theta})$ ,  $d_2(\theta, \hat{\theta})$ ,  $\hat{\theta}$  y tiempo total de simulación de 200 replicaciones. Modelo de respuesta binaria,  $n = 170$ , bajo niveles de contaminación  $\gamma_c$  y factor de contaminación de magnitud  $\gamma_b$ .

$\beta_0$	$\gamma_c$ (%)	$\gamma_b$	$ABC$	$d(\beta, \hat{\beta})$	$d_2(\beta_v, \hat{\beta}_v)$	$d(\theta, \hat{\theta})$	$d_2(\theta, \hat{\theta})$	$\hat{\theta}$	Tiempo (Seg)
T	0	1.0	0.69 (0.07)	1.83 (107.94)	33.34 (0.87)	0.44 (0.27)	0.27 (0.27)	1.58	2974.78
T	5	0.2	0.70 (0.07)	1.82 ( 90.56)	33.05 (0.95)	0.50 (0.27)	0.32 (0.29)	1.53	2643.58
T	5	0.5	0.70 (0.07)	1.83 (102.95)	33.03 (0.93)	0.56 (0.29)	0.40 (0.36)	1.45	2508.17
T	5	2.0	0.69 (0.07)	1.84 (105.96)	33.02 (0.96)	0.58 (0.28)	0.41 (0.34)	1.43	2450.44
T	5	5.0	0.68 (0.08)	1.84 (104.22)	34.03 (0.99)	0.33 (0.22)	0.16 (0.18)	1.73	2603.75
T	15	0.2	0.69 (0.08)	1.82 ( 89.64)	33.43 (0.88)	0.41 (0.27)	0.25 (0.28)	1.61	2517.69
T	15	0.5	0.70 (0.07)	1.84 (112.63)	32.86 (0.94)	0.61 (0.29)	0.45 (0.36)	1.41	2459.39
T	15	2.0	0.69 (0.07)	1.86 (109.63)	32.98 (0.94)	0.70 (0.30)	0.58 (0.43)	1.30	2546.75
T	15	5.0	0.66 (0.07)	1.82 (107.21)	34.74 (0.92)	0.24 (0.16)	0.08 (0.11)	1.97	2487.08
T	30	0.2	0.68 (0.08)	1.83 ( 90.06)	33.83 (0.88)	0.34 (0.24)	0.17 (0.22)	1.73	2482.94
T	30	0.5	0.70 (0.06)	1.85 (116.98)	32.83 (1.07)	0.73 (0.28)	0.61 (0.42)	1.27	2624.36
T	30	2.0	0.70 (0.08)	1.87 (120.16)	32.69 (0.97)	0.90 (0.31)	0.91 (0.58)	1.10	2545.89
T	30	5.0	0.68 (0.07)	1.82 (103.91)	33.02 (0.91)	0.53 (0.27)	0.35 (0.33)	1.49	2422.66
Triangulo	0	1.0	0.63 (0.08)	1.78 (120.66)	27.57 (0.92)	0.23 (0.18)	0.09 (0.12)	2.04	2909.62
Triangulo	5	0.2	0.63 (0.08)	1.78 (109.91)	27.45 (0.89)	0.23 (0.17)	0.08 (0.11)	2.02	2478.31
Triangulo	5	0.5	0.63 (0.08)	1.80 (121.34)	27.41 (1.05)	0.26 (0.18)	0.10 (0.13)	1.94	2539.55
Triangulo	5	2.0	0.64 (0.08)	1.79 (125.13)	27.38 (0.99)	0.28 (0.20)	0.12 (0.16)	1.86	2653.58
Triangulo	5	5.0	0.63 (0.09)	1.76 (143.91)	27.67 (1.09)	0.24 (0.19)	0.10 (0.13)	2.12	2673.44
Triangulo	15	0.2	0.63 (0.08)	1.77 (128.45)	27.53 (1.05)	0.25 (0.19)	0.10 (0.15)	2.03	2605.39
Triangulo	15	0.5	0.63 (0.08)	1.80 (116.56)	27.41 (0.97)	0.26 (0.19)	0.11 (0.15)	1.84	2558.64
Triangulo	15	2.0	0.64 (0.08)	1.84 (106.28)	27.58 (0.90)	0.34 (0.22)	0.17 (0.18)	1.71	2519.88
Triangulo	15	5.0	0.61 (0.08)	1.72 (166.85)	28.08 (1.25)	0.30 (0.22)	0.14 (0.18)	2.24	2449.44
Triangulo	30	0.2	0.63 (0.08)	1.77 (155.56)	27.88 (1.18)	0.27 (0.19)	0.11 (0.15)	2.12	2443.45
Triangulo	30	0.5	0.64 (0.09)	1.83 ( 99.16)	27.39 (0.92)	0.35 (0.25)	0.18 (0.21)	1.69	2596.95
Triangulo	30	2.0	0.64 (0.09)	1.86 ( 84.26)	27.49 (0.81)	0.54 (0.30)	0.38 (0.34)	1.47	2532.50
Triangulo	30	5.0	0.63 (0.08)	1.79 (117.51)	27.41 (0.96)	0.26 (0.18)	0.10 (0.13)	1.96	2442.81
Cuadrado	0	1.0	0.68 (0.08)	1.78 ( 99.06)	30.67 (0.94)	0.32 (0.23)	0.16 (0.21)	1.75	2789.80
Cuadrado	5	0.2	0.69 (0.08)	1.77 (103.66)	30.46 (1.00)	0.36 (0.23)	0.18 (0.21)	1.69	2466.44
Cuadrado	5	0.5	0.69 (0.08)	1.76 (106.28)	30.13 (0.97)	0.43 (0.28)	0.26 (0.29)	1.59	2441.83
Cuadrado	5	2.0	0.69 (0.07)	1.76 ( 91.42)	30.14 (0.89)	0.44 (0.26)	0.26 (0.29)	1.59	2440.50
Cuadrado	5	5.0	0.66 (0.08)	1.78 (104.24)	31.23 (0.91)	0.28 (0.20)	0.12 (0.16)	1.83	2677.55
Cuadrado	15	0.2	0.67 (0.07)	1.76 (101.36)	30.69 (1.01)	0.32 (0.22)	0.15 (0.17)	1.74	2496.31
Cuadrado	15	0.5	0.69 (0.08)	1.78 ( 90.14)	30.18 (0.91)	0.47 (0.29)	0.30 (0.31)	1.55	2417.39
Cuadrado	15	2.0	0.70 (0.08)	1.80 (100.77)	30.15 (0.92)	0.64 (0.27)	0.48 (0.37)	1.37	2421.36
Cuadrado	15	5.0	0.65 (0.08)	1.75 (117.48)	31.88 (0.97)	0.24 (0.17)	0.09 (0.11)	2.04	2477.77
Cuadrado	30	0.2	0.68 (0.07)	1.75 ( 94.94)	31.00 (0.85)	0.26 (0.18)	0.10 (0.14)	1.87	2578.80
Cuadrado	30	0.5	0.70 (0.07)	1.79 (107.59)	30.11 (0.92)	0.60 (0.27)	0.43 (0.34)	1.41	2463.64
Cuadrado	30	2.0	0.70 (0.07)	1.81 (107.19)	30.04 (1.03)	0.78 (0.29)	0.70 (0.48)	1.22	2422.45
Cuadrado	30	5.0	0.67 (0.07)	1.77 ( 96.32)	30.34 (0.89)	0.37 (0.22)	0.19 (0.19)	1.68	2394.73
Ajedrez	0	1.0	0.70 (0.07)	1.79 ( 96.03)	39.83 (0.87)	0.69 (0.29)	0.56 (0.43)	1.31	2898.64
Ajedrez	5	0.2	0.71 (0.06)	1.79 (108.88)	39.69 (0.96)	0.75 (0.31)	0.66 (0.49)	1.26	2536.88
Ajedrez	5	0.5	0.72 (0.07)	1.79 (105.84)	39.43 (0.89)	0.85 (0.29)	0.80 (0.50)	1.15	2439.70
Ajedrez	5	2.0	0.71 (0.07)	1.81 (129.27)	39.40 (0.92)	0.85 (0.27)	0.80 (0.46)	1.15	2443.89
Ajedrez	5	5.0	0.70 (0.08)	1.81 (102.22)	40.58 (0.98)	0.53 (0.27)	0.35 (0.31)	1.49	2479.78
Ajedrez	15	0.2	0.71 (0.07)	1.80 ( 94.72)	39.86 (0.98)	0.73 (0.29)	0.62 (0.45)	1.27	2566.58
Ajedrez	15	0.5	0.72 (0.07)	1.80 (123.46)	39.41 (0.97)	0.86 (0.28)	0.81 (0.47)	1.14	2669.53
Ajedrez	15	2.0	0.71 (0.07)	1.82 (135.20)	39.48 (0.99)	0.97 (0.26)	1.00 (0.51)	1.03	2573.83
Ajedrez	15	5.0	0.68 (0.07)	1.80 (109.36)	41.65 (0.98)	0.34 (0.23)	0.17 (0.20)	1.71	2485.42
Ajedrez	30	0.2	0.70 (0.07)	1.79 ( 94.63)	40.25 (0.92)	0.56 (0.28)	0.39 (0.33)	1.44	2413.28
Ajedrez	30	0.5	0.71 (0.08)	1.81 (124.07)	39.27 (0.96)	0.97 (0.32)	1.04 (0.62)	1.03	2434.27
Ajedrez	30	2.0	0.72 (0.06)	1.81 (115.94)	39.24 (0.90)	1.11 (0.29)	1.31 (0.65)	0.89	2427.69
Ajedrez	30	5.0	0.72 (0.07)	1.79 (130.05)	39.37 (0.91)	0.78 (0.27)	0.68 (0.44)	1.22	2402.89

**Tabla A-6.:** Promedio y desviación estándar (en paréntesis) de  $ABC$ ,  $d(\beta, \hat{\beta})$ ,  $d_2(\beta_v, \hat{\beta}_v)$ ,  $d(\theta, \hat{\theta})$ ,  $d_2(\theta, \hat{\theta})$ ,  $\hat{\theta}$  y tiempo total de simulación de 200 replicaciones. Modelo de respuesta binaria,  $n = 220$ , bajo niveles de contaminación  $\gamma_c$  y factor de contaminación de magnitud  $\gamma_b$ .



$\beta_0$	$\beta_1$	$\gamma_c$ (%)	$RMSPE$	$d(\beta, \hat{\beta})$	$d_2(\beta_v, \hat{\beta}_v)$	$d(\theta, \hat{\theta})$	$d_2(\theta, \hat{\theta})$	$\hat{\theta}$	Tiempo (Seg)
T	Tr	5.0	3.20 (0.41)	1.19 (67.33)	30.31 (0.83)	0.33 (0.25)	0.17 (0.24)	1.77	939.70
T	Tr	15.0	3.28 (0.43)	1.18 (64.05)	30.87 (0.79)	0.34 (0.23)	0.17 (0.22)	1.75	977.66
T	Tr	30.0	3.27 (0.45)	1.18 (54.62)	31.22 (0.78)	0.31 (0.20)	0.14 (0.16)	1.80	993.92
T	C	5.0	3.15 (0.41)	1.20 (64.38)	30.21 (0.81)	0.33 (0.23)	0.16 (0.20)	1.76	994.95
T	C	15.0	3.20 (0.38)	1.21 (61.61)	30.67 (0.78)	0.32 (0.22)	0.15 (0.19)	1.79	1011.00
T	C	30.0	3.29 (0.45)	1.22 (67.92)	31.20 (0.75)	0.34 (0.26)	0.18 (0.26)	1.76	1082.98
T	A	5.0	3.20 (0.42)	1.22 (75.86)	30.49 (0.85)	0.28 (0.22)	0.13 (0.17)	1.77	1076.80
T	A	15.0	3.25 (0.41)	1.28 (77.75)	31.32 (0.95)	0.36 (0.25)	0.19 (0.25)	1.74	1091.55
T	A	30.0	3.47 (0.47)	1.35 (84.77)	32.32 (0.90)	0.35 (0.23)	0.17 (0.20)	1.80	1086.64
Triangulo	T	5.0	2.30 (0.30)	0.95 (46.83)	20.91 (0.66)	0.24 (0.16)	0.08 (0.10)	1.81	978.06
Triangulo	T	15.0	2.38 (0.31)	0.99 (57.20)	21.48 (0.67)	0.28 (0.20)	0.12 (0.14)	1.78	1004.06
Triangulo	T	30.0	2.42 (0.33)	1.05 (58.09)	22.11 (0.71)	0.25 (0.18)	0.10 (0.12)	1.83	1031.59
Triangulo	C	5.0	2.35 (0.33)	0.93 (45.80)	20.85 (0.62)	0.22 (0.16)	0.08 (0.11)	1.85	1038.69
Triangulo	C	15.0	2.34 (0.32)	0.96 (55.75)	21.17 (0.68)	0.26 (0.18)	0.10 (0.12)	1.78	1051.95
Triangulo	C	30.0	2.38 (0.31)	1.00 (53.52)	21.60 (0.70)	0.26 (0.18)	0.10 (0.12)	1.83	1086.44
Triangulo	A	5.0	2.33 (0.30)	0.97 (60.02)	21.30 (0.70)	0.24 (0.17)	0.09 (0.11)	1.83	1087.27
Triangulo	A	15.0	2.52 (0.30)	1.08 (87.02)	22.54 (0.88)	0.29 (0.19)	0.12 (0.14)	1.79	1082.55
Triangulo	A	30.0	2.60 (0.33)	1.19 (87.70)	24.16 (0.86)	0.30 (0.22)	0.14 (0.17)	1.80	1023.47
Cuadrado	T	5.0	2.87 (0.38)	1.06 (60.52)	26.71 (0.73)	0.29 (0.22)	0.13 (0.17)	1.80	959.58
Cuadrado	T	15.0	2.89 (0.41)	1.08 (58.30)	27.27 (0.79)	0.29 (0.22)	0.13 (0.17)	1.80	956.81
Cuadrado	T	30.0	2.99 (0.40)	1.12 (58.79)	27.91 (0.79)	0.30 (0.23)	0.14 (0.19)	1.77	986.36
Cuadrado	Tr	5.0	2.83 (0.37)	1.05 (54.46)	26.75 (0.73)	0.27 (0.20)	0.11 (0.16)	1.83	1021.64
Cuadrado	Tr	15.0	2.87 (0.40)	1.04 (57.09)	27.13 (0.66)	0.28 (0.20)	0.12 (0.15)	1.79	1029.38
Cuadrado	Tr	30.0	2.89 (0.36)	1.04 (51.89)	27.46 (0.71)	0.29 (0.21)	0.13 (0.17)	1.80	1027.16
Cuadrado	A	5.0	2.87 (0.39)	1.09 (61.31)	27.10 (0.76)	0.28 (0.21)	0.13 (0.16)	1.79	1051.34
Cuadrado	A	15.0	2.97 (0.38)	1.16 (74.11)	27.88 (0.86)	0.32 (0.23)	0.16 (0.20)	1.77	1057.59
Cuadrado	A	30.0	3.12 (0.39)	1.24 (86.93)	29.20 (0.97)	0.32 (0.24)	0.16 (0.22)	1.78	1064.69
Ajedrez	T	5.0	3.93 (0.56)	1.37 (82.63)	37.71 (1.07)	0.36 (0.27)	0.20 (0.28)	1.72	931.89
Ajedrez	T	15.0	3.98 (0.49)	1.38 (83.23)	38.51 (1.11)	0.37 (0.29)	0.22 (0.31)	1.75	1031.36
Ajedrez	T	30.0	4.02 (0.58)	1.39 (84.00)	39.48 (1.13)	0.37 (0.27)	0.21 (0.25)	1.76	1051.94
Ajedrez	Tr	5.0	3.90 (0.54)	1.37 (81.04)	37.82 (1.09)	0.38 (0.26)	0.21 (0.24)	1.76	1046.59
Ajedrez	Tr	15.0	4.00 (0.50)	1.36 (87.12)	38.53 (1.07)	0.36 (0.26)	0.20 (0.26)	1.76	1047.80
Ajedrez	Tr	30.0	4.12 (0.57)	1.33 (86.86)	39.85 (1.05)	0.38 (0.26)	0.21 (0.24)	1.72	1054.16
Ajedrez	C	5.0	3.88 (0.48)	1.38 (85.29)	37.76 (1.08)	0.36 (0.27)	0.20 (0.28)	1.75	1054.27
Ajedrez	C	15.0	3.99 (0.56)	1.37 (88.94)	38.36 (1.01)	0.36 (0.26)	0.20 (0.26)	1.75	1055.06
Ajedrez	C	30.0	4.09 (0.53)	1.38 (88.61)	39.44 (1.00)	0.36 (0.25)	0.19 (0.24)	1.75	1026.44

**Tabla A-7.:** Promedio y desviación estándar (en paréntesis) de  $RMSPE$ ,  $d(\beta, \hat{\beta})$ ,  $d_2(\beta_v, \hat{\beta}_v)$ ,  $d(\theta, \hat{\theta})$ ,  $d_2(\theta, \hat{\theta})$ ,  $\hat{\theta}$  y tiempo total de simulación de 200 replicaciones. Modelo de respuesta continua,  $n = 100$ , bajo niveles de contaminación  $\gamma_c$  y contaminación de cambio de forma - Tipo 1 en parámetro imagen.

$\beta_0$	$\beta_1$	$\gamma_c$	$RMSPE$	$d(\beta, \hat{\beta})$	$d_2(\beta_v, \hat{\beta}_v)$	$d(\theta, \hat{\theta})$	$d_2(\theta, \hat{\theta})$	$\hat{\theta}$	tiempo
T	Tr	5.0	2.93 (0.29)	1.27 (47.46)	27.83 (0.73)	0.21 (0.15)	0.07 (0.10)	1.86	1359.66
T	Tr	15.0	3.04 (0.29)	1.27 (57.79)	28.51 (0.78)	0.22 (0.16)	0.07 (0.11)	1.85	1320.48
T	Tr	30.0	3.14 (0.31)	1.27 (54.19)	29.45 (0.71)	0.21 (0.15)	0.07 (0.08)	1.88	1322.02
T	C	5.0	2.95 (0.30)	1.27 (50.09)	27.66 (0.85)	0.22 (0.16)	0.07 (0.09)	1.88	1306.78
T	C	15.0	3.01 (0.31)	1.30 (50.55)	28.51 (0.81)	0.20 (0.15)	0.06 (0.08)	1.87	1319.23
T	C	30.0	3.10 (0.32)	1.31 (56.48)	29.39 (0.80)	0.21 (0.16)	0.07 (0.09)	1.89	1351.27
T	A	5.0	3.01 (0.31)	1.30 (59.04)	28.24 (0.87)	0.24 (0.18)	0.09 (0.13)	1.86	1323.58
T	A	15.0	3.10 (0.32)	1.38 (70.34)	29.38 (0.91)	0.23 (0.17)	0.08 (0.11)	1.87	1334.39
T	A	30.0	3.24 (0.34)	1.47 (71.16)	31.04 (0.87)	0.26 (0.19)	0.10 (0.13)	1.87	1360.30
Triangulo	T	5.0	2.17 (0.23)	1.00 (46.77)	19.17 (0.64)	0.14 (0.10)	0.03 (0.04)	1.90	1081.05
Triangulo	T	15.0	2.25 (0.24)	1.08 (56.90)	20.13 (0.73)	0.18 (0.15)	0.05 (0.08)	1.88	1080.62
Triangulo	T	30.0	2.34 (0.24)	1.14 (58.09)	21.19 (0.73)	0.19 (0.14)	0.06 (0.08)	1.89	1091.25
Triangulo	C	5.0	2.13 (0.21)	0.99 (44.56)	18.92 (0.64)	0.16 (0.12)	0.04 (0.05)	1.90	1072.72
Triangulo	C	15.0	2.19 (0.23)	1.03 (46.81)	19.45 (0.70)	0.17 (0.12)	0.04 (0.06)	1.90	1237.39
Triangulo	C	30.0	2.28 (0.22)	1.08 (50.29)	20.23 (0.63)	0.18 (0.14)	0.05 (0.07)	1.87	1310.08
Triangulo	A	5.0	2.20 (0.23)	1.04 (56.18)	19.65 (0.72)	0.17 (0.12)	0.05 (0.06)	1.87	1316.56
Triangulo	A	15.0	2.40 (0.26)	1.17 (66.11)	21.62 (0.86)	0.20 (0.15)	0.06 (0.09)	1.89	1303.58
Triangulo	A	30.0	2.58 (0.25)	1.33 (69.92)	24.01 (0.96)	0.20 (0.15)	0.06 (0.09)	1.90	1310.42
Cuadrado	T	5.0	2.65 (0.28)	1.13 (42.83)	24.33 (0.72)	0.20 (0.14)	0.06 (0.07)	1.88	1255.61
Cuadrado	T	15.0	2.65 (0.28)	1.18 (51.92)	25.09 (0.75)	0.19 (0.15)	0.06 (0.09)	1.87	1298.09
Cuadrado	T	30.0	2.81 (0.30)	1.23 (57.78)	26.15 (0.83)	0.22 (0.16)	0.07 (0.10)	1.83	1301.00
Cuadrado	Tr	5.0	2.59 (0.26)	1.11 (44.53)	24.12 (0.71)	0.19 (0.14)	0.06 (0.07)	1.88	1290.55
Cuadrado	Tr	15.0	2.64 (0.28)	1.12 (45.94)	24.63 (0.71)	0.19 (0.14)	0.05 (0.07)	1.87	1290.05
Cuadrado	Tr	30.0	2.74 (0.29)	1.12 (51.27)	25.24 (0.79)	0.20 (0.14)	0.06 (0.08)	1.88	1288.69
Cuadrado	A	5.0	2.67 (0.24)	1.16 (52.77)	24.72 (0.78)	0.18 (0.13)	0.05 (0.07)	1.90	1290.20
Cuadrado	A	15.0	2.77 (0.29)	1.27 (65.60)	26.22 (0.92)	0.21 (0.16)	0.07 (0.10)	1.87	1290.34
Cuadrado	A	30.0	2.99 (0.30)	1.37 (77.93)	28.20 (1.03)	0.23 (0.17)	0.08 (0.11)	1.88	1280.78
Ajedrez	T	5.0	3.59 (0.37)	1.45 (62.94)	34.64 (1.10)	0.25 (0.18)	0.10 (0.12)	1.84	1390.53
Ajedrez	T	15.0	3.68 (0.36)	1.48 (64.85)	35.68 (1.12)	0.24 (0.17)	0.09 (0.11)	1.87	1582.56
Ajedrez	T	30.0	3.82 (0.39)	1.50 (70.83)	37.40 (1.11)	0.26 (0.19)	0.11 (0.14)	1.84	1569.70
Ajedrez	Tr	5.0	3.60 (0.35)	1.44 (67.99)	34.75 (1.18)	0.23 (0.16)	0.08 (0.10)	1.86	1610.45
Ajedrez	Tr	15.0	3.71 (0.38)	1.46 (69.41)	35.95 (1.09)	0.25 (0.19)	0.10 (0.15)	1.87	1641.22
Ajedrez	Tr	30.0	3.92 (0.42)	1.45 (67.28)	37.61 (1.05)	0.24 (0.18)	0.09 (0.11)	1.85	1541.92
Ajedrez	C	5.0	3.68 (0.37)	1.45 (65.61)	34.78 (1.16)	0.28 (0.19)	0.11 (0.14)	1.83	1467.94
Ajedrez	C	15.0	3.74 (0.39)	1.46 (65.64)	35.93 (1.19)	0.24 (0.18)	0.09 (0.14)	1.86	1489.31
Ajedrez	C	30.0	3.86 (0.42)	1.48 (73.76)	37.51 (1.01)	0.26 (0.19)	0.11 (0.14)	1.87	1549.23

**Tabla A-8.:** Promedio y desviación estándar (en paréntesis) de  $RMSPE$ ,  $d(\beta, \hat{\beta})$ ,  $d_2(\beta_v, \hat{\beta}_v)$ ,  $d(\theta, \hat{\theta})$ ,  $d_2(\theta, \hat{\theta})$ ,  $\hat{\theta}$  y tiempo total de simulación de 200 replicaciones. Modelo de respuesta continua,  $n = 170$ , bajo niveles de contaminación  $\gamma_c$  y contaminación de cambio de forma - Tipo 1 en parámetro imagen.

$\beta_0$	$\beta_1$	$\gamma_c$	$RMSPE$	$d(\beta, \hat{\beta})$	$d_2(\beta_v, \hat{\beta}_v)$	$d(\theta, \hat{\theta})$	$d_2(\theta, \hat{\theta})$	$\hat{\theta}$	tiempo
T	Tr	5.0	2.79 (0.26)	1.29 (48.94)	26.26 (0.83)	0.17 (0.13)	0.05 (0.06)	1.89	1722.02
T	Tr	15.0	2.91 (0.26)	1.31 (46.88)	27.24 (0.75)	0.18 (0.14)	0.05 (0.07)	1.88	1610.41
T	Tr	30.0	2.98 (0.31)	1.33 (50.79)	28.34 (0.79)	0.17 (0.14)	0.05 (0.07)	1.90	1627.52
T	C	5.0	2.81 (0.26)	1.30 (49.67)	26.34 (0.68)	0.16 (0.12)	0.04 (0.05)	1.91	1631.78
T	C	15.0	2.90 (0.27)	1.34 (50.65)	27.22 (0.81)	0.19 (0.14)	0.05 (0.08)	1.90	1632.50
T	C	30.0	2.99 (0.25)	1.36 (53.14)	28.21 (0.82)	0.15 (0.11)	0.04 (0.05)	1.92	1630.77
T	A	5.0	2.88 (0.28)	1.35 (55.33)	26.77 (0.79)	0.18 (0.13)	0.05 (0.07)	1.90	1597.53
T	A	15.0	3.03 (0.27)	1.44 (67.42)	28.33 (0.89)	0.20 (0.15)	0.06 (0.09)	1.90	1624.06
T	A	30.0	3.18 (0.29)	1.54 (64.61)	30.25 (1.03)	0.21 (0.14)	0.06 (0.07)	1.90	1634.64
Triangulo	T	5.0	2.08 (0.20)	1.03 (44.03)	18.15 (0.66)	0.13 (0.10)	0.03 (0.04)	1.92	1649.52
Triangulo	T	15.0	2.20 (0.19)	1.13 (53.72)	19.45 (0.69)	0.14 (0.09)	0.03 (0.04)	1.93	1664.34
Triangulo	T	30.0	2.33 (0.20)	1.21 (57.61)	20.74 (0.77)	0.16 (0.11)	0.04 (0.05)	1.91	1712.67
Triangulo	C	5.0	2.04 (0.18)	1.01 (42.08)	17.76 (0.60)	0.12 (0.09)	0.02 (0.03)	1.92	1719.98
Triangulo	C	15.0	2.10 (0.18)	1.06 (48.17)	18.56 (0.70)	0.14 (0.11)	0.03 (0.05)	1.91	1776.67
Triangulo	C	30.0	2.20 (0.21)	1.12 (47.53)	19.50 (0.65)	0.14 (0.12)	0.03 (0.05)	1.90	1734.47
Triangulo	A	5.0	2.17 (0.21)	1.10 (68.35)	19.09 (0.87)	0.15 (0.11)	0.03 (0.05)	1.92	1716.78
Triangulo	A	15.0	2.37 (0.23)	1.25 (68.89)	21.32 (1.02)	0.19 (0.13)	0.05 (0.06)	1.89	1693.52
Triangulo	A	30.0	2.59 (0.23)	1.41 (69.18)	23.94 (0.97)	0.19 (0.14)	0.06 (0.07)	1.90	1727.05
Cuadrado	T	5.0	2.49 (0.22)	1.16 (44.79)	22.78 (0.76)	0.16 (0.12)	0.04 (0.05)	1.89	1642.28
Cuadrado	T	15.0	2.61 (0.24)	1.22 (52.84)	23.89 (0.82)	0.16 (0.12)	0.04 (0.06)	1.91	1657.67
Cuadrado	T	30.0	2.70 (0.26)	1.27 (58.38)	25.20 (0.82)	0.16 (0.13)	0.04 (0.07)	1.91	1729.52
Cuadrado	Tr	5.0	2.48 (0.22)	1.13 (45.03)	22.59 (0.74)	0.16 (0.13)	0.04 (0.06)	1.91	1709.06
Cuadrado	Tr	15.0	2.54 (0.24)	1.15 (45.46)	23.33 (0.76)	0.15 (0.11)	0.03 (0.04)	1.92	1695.56
Cuadrado	Tr	30.0	2.61 (0.25)	1.17 (48.30)	24.15 (0.76)	0.15 (0.11)	0.03 (0.04)	1.91	1717.53
Cuadrado	A	5.0	2.55 (0.23)	1.20 (52.34)	23.47 (0.89)	0.16 (0.12)	0.04 (0.06)	1.90	1685.94
Cuadrado	A	15.0	2.76 (0.23)	1.34 (63.95)	25.42 (0.99)	0.18 (0.14)	0.05 (0.07)	1.90	1745.98
Cuadrado	A	30.0	2.95 (0.28)	1.45 (67.06)	27.68 (0.96)	0.21 (0.15)	0.07 (0.09)	1.91	1843.55
Ajedrez	T	5.0	3.47 (0.33)	1.49 (51.54)	32.82 (1.20)	0.21 (0.15)	0.06 (0.09)	1.88	1299.73
Ajedrez	T	15.0	3.60 (0.32)	1.53 (60.61)	34.31 (1.13)	0.21 (0.16)	0.07 (0.10)	1.89	1705.06
Ajedrez	T	30.0	3.76 (0.37)	1.57 (68.31)	36.30 (1.10)	0.21 (0.14)	0.06 (0.08)	1.93	1764.19
Ajedrez	Tr	5.0	3.45 (0.28)	1.48 (59.51)	32.94 (1.18)	0.22 (0.16)	0.07 (0.11)	1.88	1722.19
Ajedrez	Tr	15.0	3.58 (0.33)	1.51 (62.20)	34.57 (1.03)	0.22 (0.17)	0.08 (0.11)	1.87	1813.33
Ajedrez	Tr	30.0	3.82 (0.32)	1.53 (64.45)	36.74 (1.08)	0.21 (0.15)	0.07 (0.08)	1.86	1868.47
Ajedrez	C	5.0	3.45 (0.33)	1.48 (55.02)	32.97 (1.17)	0.21 (0.16)	0.07 (0.10)	1.89	1870.75
Ajedrez	C	15.0	3.61 (0.34)	1.52 (60.06)	34.46 (1.06)	0.23 (0.17)	0.08 (0.11)	1.89	1868.45
Ajedrez	C	30.0	3.78 (0.34)	1.56 (62.43)	36.57 (1.07)	0.23 (0.17)	0.08 (0.11)	1.88	1865.09

**Tabla A-9.:** Promedio y desviación estándar (en paréntesis) de  $RMSPE$ ,  $d(\beta, \hat{\beta})$ ,  $d_2(\beta_v, \hat{\beta}_v)$ ,  $d(\theta, \hat{\theta})$ ,  $d_2(\theta, \hat{\theta})$ ,  $\hat{\theta}$  y tiempo total de simulación de 200 replicaciones. Modelo de respuesta continua,  $n = 220$ , bajo niveles de contaminación  $\gamma_c$  y contaminación de cambio de forma - Tipo 1 en parámetro imagen.

$\beta_0$	$\beta_1$	$\gamma_c(\%)$	$ABC$	$d(\beta, \hat{\beta})$	$d_2(\beta_v, \hat{\beta}_v)$	$d(\theta, \hat{\theta})$	$d_2(\theta, \hat{\theta})$	$\hat{\theta}$	Tiempo (Seg)
T	Tr	5.0	0.64 (0.12)	1.35 (61.34)	33.75 (0.78)	1.10 (0.35)	1.32 (0.76)	0.90	1445.98
T	Tr	15.0	0.63 (0.12)	1.35 (63.80)	33.96 (0.78)	1.05 (0.34)	1.21 (0.72)	0.95	1418.16
T	Tr	30.0	0.63 (0.12)	1.35 (66.56)	34.53 (0.86)	0.97 (0.39)	1.10 (0.79)	1.03	1152.55
T	C	5.0	0.67 (0.11)	1.34 (54.52)	33.61 (0.78)	1.04 (0.35)	1.21 (0.73)	0.96	1079.08
T	C	15.0	0.65 (0.11)	1.35 (64.66)	33.89 (0.85)	1.09 (0.36)	1.32 (0.81)	0.91	1095.11
T	C	30.0	0.62 (0.12)	1.36 (54.01)	34.33 (0.75)	1.05 (0.37)	1.23 (0.84)	0.96	1111.45
T	A	5.0	0.64 (0.10)	1.35 (56.37)	33.74 (0.74)	1.11 (0.34)	1.36 (0.77)	0.89	1105.09
T	A	15.0	0.63 (0.12)	1.36 (57.79)	34.09 (0.78)	1.09 (0.37)	1.32 (0.79)	0.91	1046.28
T	A	30.0	0.64 (0.11)	1.37 (62.84)	34.44 (0.82)	1.18 (0.36)	1.53 (0.88)	0.82	820.88
Triangulo	T	5.0	0.59 (0.13)	1.31 (64.41)	25.90 (0.69)	0.70 (0.35)	0.61 (0.53)	1.32	799.88
Triangulo	T	15.0	0.60 (0.11)	1.32 (62.61)	26.05 (0.67)	0.76 (0.34)	0.69 (0.55)	1.25	782.09
Triangulo	T	30.0	0.58 (0.14)	1.32 (58.24)	26.22 (0.69)	0.78 (0.34)	0.73 (0.57)	1.22	715.64
Triangulo	C	5.0	0.61 (0.13)	1.31 (71.25)	25.76 (0.69)	0.68 (0.36)	0.59 (0.54)	1.33	752.62
Triangulo	C	15.0	0.60 (0.14)	1.31 (66.99)	25.83 (0.72)	0.73 (0.37)	0.67 (0.60)	1.28	755.52
Triangulo	C	30.0	0.61 (0.13)	1.31 (68.10)	25.89 (0.65)	0.69 (0.36)	0.60 (0.53)	1.32	757.73
Triangulo	A	5.0	0.60 (0.12)	1.32 (60.82)	26.00 (0.68)	0.69 (0.31)	0.58 (0.45)	1.31	760.16
Triangulo	A	15.0	0.57 (0.13)	1.33 (65.80)	26.32 (0.72)	0.75 (0.38)	0.71 (0.60)	1.25	759.05
Triangulo	A	30.0	0.59 (0.13)	1.35 (59.73)	26.68 (0.72)	0.88 (0.36)	0.90 (0.67)	1.12	777.84
Cuadrado	T	5.0	0.63 (0.12)	1.28 (62.20)	30.35 (0.83)	0.91 (0.36)	0.95 (0.67)	1.09	821.28
Cuadrado	T	15.0	0.64 (0.12)	1.30 (59.32)	30.67 (0.83)	0.93 (0.37)	1.00 (0.70)	1.07	816.53
Cuadrado	T	30.0	0.64 (0.12)	1.30 (64.65)	30.94 (0.81)	0.98 (0.41)	1.12 (0.82)	1.03	802.42
Cuadrado	Tr	5.0	0.66 (0.13)	1.28 (57.57)	30.39 (0.79)	0.94 (0.35)	1.01 (0.68)	1.06	771.00
Cuadrado	Tr	15.0	0.63 (0.11)	1.29 (59.61)	30.53 (0.78)	0.90 (0.35)	0.93 (0.67)	1.10	768.66
Cuadrado	Tr	30.0	0.63 (0.12)	1.29 (65.60)	30.89 (0.85)	0.87 (0.36)	0.89 (0.64)	1.13	768.34
Cuadrado	A	5.0	0.64 (0.11)	1.28 (57.93)	30.49 (0.80)	0.92 (0.34)	0.96 (0.62)	1.08	766.45
Cuadrado	A	15.0	0.64 (0.12)	1.30 (66.47)	30.93 (0.84)	0.99 (0.37)	1.11 (0.75)	1.01	762.67
Cuadrado	A	30.0	0.62 (0.12)	1.32 (62.12)	31.44 (0.85)	1.08 (0.35)	1.28 (0.78)	0.92	767.92
Ajedrez	T	5.0	0.67 (0.11)	1.30 (55.04)	41.63 (0.72)	1.35 (0.33)	1.94 (0.89)	0.65	1184.36
Ajedrez	T	15.0	0.66 (0.11)	1.31 (58.10)	41.96 (0.77)	1.25 (0.34)	1.67 (0.82)	0.75	1555.62
Ajedrez	T	30.0	0.64 (0.12)	1.33 (55.72)	42.74 (0.88)	1.17 (0.34)	1.48 (0.79)	0.83	1605.31
Ajedrez	Tr	5.0	0.66 (0.10)	1.30 (58.39)	41.62 (0.81)	1.24 (0.32)	1.65 (0.79)	0.76	1514.42
Ajedrez	Tr	15.0	0.65 (0.12)	1.31 (57.68)	42.19 (0.86)	1.16 (0.33)	1.47 (0.79)	0.84	1519.73
Ajedrez	Tr	30.0	0.64 (0.12)	1.32 (60.94)	42.93 (0.87)	1.09 (0.33)	1.30 (0.72)	0.91	1516.84
Ajedrez	C	5.0	0.67 (0.11)	1.30 (53.54)	41.57 (0.80)	1.30 (0.31)	1.78 (0.82)	0.70	1465.12
Ajedrez	C	15.0	0.65 (0.11)	1.32 (54.54)	42.14 (0.87)	1.24 (0.33)	1.64 (0.79)	0.76	1470.88
Ajedrez	C	30.0	0.64 (0.11)	1.32 (65.60)	42.75 (0.85)	1.18 (0.32)	1.49 (0.75)	0.82	1386.33

**Tabla A-10.:** Promedio y desviación estándar (en paréntesis) de  $ABC$ ,  $d(\beta, \hat{\beta})$ ,  $d_2(\beta_v, \hat{\beta}_v)$ ,  $d(\theta, \hat{\theta})$ ,  $d_2(\theta, \hat{\theta})$ ,  $\hat{\theta}$  y tiempo total de simulación de 200 replicaciones. Modelo de respuesta binaria,  $n = 100$ , bajo niveles de contaminación  $\gamma_c$  y contaminación de cambio de forma - Tipo 1 en parámetro imagen.

$\beta_0$	$\beta_1$	$\gamma_c$ (%)	$ABC$	$d(\beta, \hat{\beta})$	$d_2(\beta_v, \hat{\beta}_v)$	$d(\theta, \hat{\theta})$	$d_2(\theta, \hat{\theta})$	$\hat{\theta}$	Tiempo (Seg)
T	Tr	5.0	0.67 (0.09)	1.53 (134.90)	32.93 (0.84)	0.80 (0.29)	0.73 (0.46)	1.20	1624.98
T	Tr	15.0	0.67 (0.08)	1.54 (133.71)	33.40 (0.86)	0.78 (0.29)	0.70 (0.45)	1.22	1709.25
T	Tr	30.0	0.65 (0.08)	1.56 (141.63)	34.09 (0.87)	0.68 (0.29)	0.55 (0.39)	1.32	1455.98
T	C	5.0	0.68 (0.08)	1.53 (140.59)	32.92 (0.89)	0.85 (0.29)	0.81 (0.48)	1.15	1476.84
T	C	15.0	0.66 (0.08)	1.54 (144.42)	33.28 (0.93)	0.81 (0.28)	0.74 (0.47)	1.19	1465.84
T	C	30.0	0.67 (0.08)	1.56 (154.59)	33.80 (0.98)	0.82 (0.30)	0.77 (0.48)	1.18	1505.27
T	A	5.0	0.69 (0.08)	1.56 (153.31)	32.98 (0.96)	0.87 (0.31)	0.86 (0.54)	1.13	1493.33
T	A	15.0	0.67 (0.08)	1.56 (150.79)	33.45 (0.90)	0.87 (0.31)	0.86 (0.53)	1.13	1467.69
T	A	30.0	0.64 (0.09)	1.61 (151.08)	34.56 (1.06)	0.89 (0.29)	0.87 (0.50)	1.11	1507.53
Triangulo	T	5.0	0.62 (0.09)	1.48 (116.03)	25.88 (0.96)	0.52 (0.27)	0.35 (0.33)	1.49	1506.06
Triangulo	T	15.0	0.61 (0.09)	1.48 (113.69)	26.18 (0.95)	0.54 (0.28)	0.37 (0.32)	1.47	1306.61
Triangulo	T	30.0	0.61 (0.08)	1.51 (132.04)	26.50 (1.07)	0.59 (0.27)	0.42 (0.33)	1.42	1267.69
Triangulo	C	5.0	0.63 (0.09)	1.48 (123.27)	25.92 (0.99)	0.50 (0.26)	0.32 (0.30)	1.51	1272.88
Triangulo	C	15.0	0.62 (0.10)	1.48 (123.40)	26.01 (0.94)	0.49 (0.26)	0.31 (0.28)	1.52	1324.92
Triangulo	C	30.0	0.62 (0.10)	1.48 (118.88)	25.98 (0.92)	0.58 (0.27)	0.41 (0.33)	1.43	1318.55
Triangulo	A	5.0	0.62 (0.10)	1.48 (107.35)	26.02 (0.81)	0.52 (0.27)	0.34 (0.30)	1.49	1324.28
Triangulo	A	15.0	0.62 (0.09)	1.50 (120.58)	26.56 (0.99)	0.55 (0.27)	0.37 (0.33)	1.47	1299.84
Triangulo	A	30.0	0.60 (0.08)	1.54 (124.34)	27.19 (1.08)	0.66 (0.28)	0.51 (0.38)	1.35	1316.33
Cuadrado	T	5.0	0.67 (0.08)	1.47 (142.88)	29.81 (1.05)	0.73 (0.29)	0.62 (0.44)	1.27	2035.88
Cuadrado	T	15.0	0.67 (0.09)	1.49 (148.82)	30.24 (1.02)	0.75 (0.30)	0.65 (0.44)	1.25	2234.67
Cuadrado	T	30.0	0.66 (0.08)	1.51 (145.85)	30.83 (1.01)	0.76 (0.28)	0.65 (0.45)	1.24	2198.14
Cuadrado	Tr	5.0	0.68 (0.09)	1.46 (135.88)	29.67 (0.92)	0.71 (0.31)	0.60 (0.45)	1.29	2106.23
Cuadrado	Tr	15.0	0.67 (0.09)	1.46 (133.35)	29.98 (0.97)	0.69 (0.30)	0.57 (0.44)	1.31	2042.69
Cuadrado	Tr	30.0	0.65 (0.09)	1.47 (130.00)	30.52 (0.90)	0.68 (0.30)	0.55 (0.42)	1.33	2149.86
Cuadrado	A	5.0	0.68 (0.08)	1.49 (141.76)	29.98 (0.86)	0.75 (0.29)	0.64 (0.44)	1.25	2077.72
Cuadrado	A	15.0	0.66 (0.08)	1.50 (153.74)	30.58 (1.12)	0.78 (0.29)	0.69 (0.46)	1.22	1961.39
Cuadrado	A	30.0	0.64 (0.08)	1.53 (156.30)	31.36 (1.14)	0.85 (0.30)	0.81 (0.53)	1.15	1498.88
Ajedrez	T	5.0	0.71 (0.08)	1.51 (147.45)	40.27 (0.79)	1.03 (0.32)	1.16 (0.66)	0.97	2358.89
Ajedrez	T	15.0	0.69 (0.08)	1.53 (154.23)	40.89 (0.92)	1.03 (0.30)	1.15 (0.58)	0.97	2229.02
Ajedrez	T	30.0	0.67 (0.08)	1.55 (159.13)	41.89 (0.91)	1.00 (0.31)	1.10 (0.64)	1.00	2111.66
Ajedrez	Tr	5.0	0.71 (0.08)	1.51 (152.21)	40.37 (0.80)	1.02 (0.32)	1.15 (0.63)	0.98	2069.89
Ajedrez	Tr	15.0	0.68 (0.08)	1.51 (145.34)	40.99 (0.92)	0.95 (0.30)	0.99 (0.56)	1.05	2200.23
Ajedrez	Tr	30.0	0.66 (0.08)	1.53 (150.10)	42.34 (1.03)	0.89 (0.31)	0.89 (0.56)	1.11	2229.72
Ajedrez	C	5.0	0.70 (0.08)	1.52 (151.25)	40.19 (0.82)	1.02 (0.29)	1.13 (0.61)	0.98	2253.59
Ajedrez	C	15.0	0.68 (0.08)	1.53 (142.64)	41.03 (0.86)	1.02 (0.31)	1.15 (0.65)	0.98	2247.53
Ajedrez	C	30.0	0.66 (0.09)	1.54 (155.96)	41.97 (1.00)	0.94 (0.33)	0.99 (0.60)	1.06	2182.19

**Tabla A-11.:** Promedio y desviación estándar (en paréntesis) de  $ABC$ ,  $d(\beta, \hat{\beta})$ ,  $d_2(\beta_v, \hat{\beta}_v)$ ,  $d(\theta, \hat{\theta})$ ,  $d_2(\theta, \hat{\theta})$ ,  $\hat{\theta}$  y tiempo total de simulación de 200 replicaciones. Modelo de respuesta binaria,  $n = 170$ , bajo niveles de contaminación  $\gamma_c$  y contaminación de cambio de forma - Tipo 1 en parámetro imagen.

$\beta_0$	$\beta_1$	$\gamma_c$ (%)	$ABC$	$d(\beta, \hat{\beta})$	$d_2(\beta_v, \hat{\beta}_v)$	$d(\theta, \hat{\theta})$	$d_2(\theta, \hat{\theta})$	$\hat{\theta}$	Tiempo (Seg)
T	Tr	5.0	0.69 (0.07)	1.83 (106.44)	33.23 (0.95)	0.52 (0.28)	0.35 (0.35)	1.50	2598.89
T	Tr	15.0	0.69 (0.07)	1.86 (109.90)	33.91 (0.96)	0.45 (0.26)	0.27 (0.27)	1.56	2228.14
T	Tr	30.0	0.66 (0.07)	1.87 ( 95.70)	34.77 (0.91)	0.44 (0.28)	0.28 (0.30)	1.59	3268.84
T	C	5.0	0.69 (0.08)	1.84 (113.13)	33.39 (0.95)	0.48 (0.28)	0.31 (0.31)	1.53	3253.47
T	C	15.0	0.68 (0.08)	1.85 ( 88.88)	33.77 (0.96)	0.51 (0.26)	0.33 (0.28)	1.51	3259.20
T	C	30.0	0.66 (0.07)	1.88 (115.76)	34.45 (0.98)	0.49 (0.27)	0.32 (0.32)	1.52	3093.70
T	A	5.0	0.69 (0.07)	1.84 ( 91.67)	33.48 (0.94)	0.51 (0.27)	0.33 (0.30)	1.51	2633.58
T	A	15.0	0.67 (0.07)	1.87 (100.86)	34.20 (0.94)	0.55 (0.29)	0.39 (0.35)	1.46	2237.48
T	A	30.0	0.66 (0.07)	1.91 (116.65)	35.17 (1.18)	0.67 (0.30)	0.53 (0.43)	1.34	2340.73
Triangulo	T	5.0	0.62 (0.08)	1.79 (111.68)	27.57 (0.94)	0.22 (0.18)	0.08 (0.12)	1.96	2034.78
Triangulo	T	15.0	0.63 (0.08)	1.81 (116.56)	27.90 (1.05)	0.24 (0.19)	0.10 (0.14)	1.88	2026.17
Triangulo	T	30.0	0.63 (0.08)	1.83 (104.48)	28.32 (1.00)	0.29 (0.22)	0.13 (0.17)	1.80	1827.61
Triangulo	C	5.0	0.63 (0.09)	1.79 (109.65)	27.46 (0.90)	0.23 (0.18)	0.09 (0.13)	1.97	1773.34
Triangulo	C	15.0	0.64 (0.08)	1.81 (117.08)	27.67 (0.99)	0.26 (0.18)	0.10 (0.13)	1.90	1962.14
Triangulo	C	30.0	0.63 (0.08)	1.81 (120.06)	27.80 (1.08)	0.27 (0.19)	0.11 (0.14)	1.87	2130.09
Triangulo	A	5.0	0.63 (0.09)	1.81 (116.18)	27.82 (1.01)	0.23 (0.18)	0.09 (0.13)	1.92	1886.42
Triangulo	A	15.0	0.62 (0.08)	1.85 (105.75)	28.53 (0.99)	0.26 (0.20)	0.11 (0.15)	1.88	1835.03
Triangulo	A	30.0	0.63 (0.08)	1.87 (100.37)	29.31 (1.09)	0.30 (0.22)	0.14 (0.18)	1.77	1839.64
Cuadrado	T	5.0	0.68 (0.08)	1.80 (116.28)	30.63 (1.03)	0.41 (0.23)	0.22 (0.21)	1.64	2546.33
Cuadrado	T	15.0	0.68 (0.08)	1.80 ( 85.05)	31.04 (0.88)	0.39 (0.26)	0.21 (0.26)	1.67	2415.03
Cuadrado	T	30.0	0.67 (0.08)	1.82 (100.90)	31.70 (1.03)	0.44 (0.28)	0.27 (0.31)	1.59	2423.25
Cuadrado	Tr	5.0	0.68 (0.07)	1.78 ( 84.47)	30.47 (0.97)	0.35 (0.23)	0.18 (0.21)	1.68	2396.41
Cuadrado	Tr	15.0	0.67 (0.08)	1.79 ( 92.12)	30.92 (0.95)	0.37 (0.25)	0.20 (0.24)	1.68	2468.19
Cuadrado	Tr	30.0	0.66 (0.08)	1.79 ( 91.14)	31.32 (0.91)	0.33 (0.22)	0.16 (0.18)	1.72	2478.95
Cuadrado	A	5.0	0.69 (0.07)	1.79 ( 93.94)	30.77 (1.02)	0.43 (0.24)	0.24 (0.23)	1.60	2498.12
Cuadrado	A	15.0	0.66 (0.07)	1.82 (102.62)	31.41 (0.99)	0.45 (0.27)	0.27 (0.27)	1.57	2493.61
Cuadrado	A	30.0	0.65 (0.07)	1.86 (112.65)	32.58 (1.18)	0.46 (0.27)	0.29 (0.29)	1.56	2482.27
Ajedrez	T	5.0	0.70 (0.07)	1.80 (108.53)	39.84 (0.93)	0.82 (0.28)	0.75 (0.47)	1.18	1799.70
Ajedrez	T	15.0	0.70 (0.06)	1.84 (117.63)	40.69 (1.01)	0.75 (0.29)	0.64 (0.47)	1.25	1674.05
Ajedrez	T	30.0	0.68 (0.07)	1.87 (109.63)	41.94 (1.09)	0.73 (0.30)	0.62 (0.45)	1.27	1675.83
Ajedrez	Tr	5.0	0.71 (0.06)	1.81 ( 99.27)	39.98 (0.91)	0.76 (0.30)	0.66 (0.50)	1.25	1664.89
Ajedrez	Tr	15.0	0.69 (0.08)	1.83 ( 92.03)	41.01 (0.94)	0.67 (0.29)	0.54 (0.42)	1.33	1664.67
Ajedrez	Tr	30.0	0.68 (0.07)	1.88 (112.58)	42.46 (1.01)	0.58 (0.27)	0.41 (0.32)	1.43	1676.27
Ajedrez	C	5.0	0.71 (0.07)	1.81 (112.34)	39.86 (0.94)	0.77 (0.27)	0.67 (0.40)	1.23	1679.53
Ajedrez	C	15.0	0.69 (0.08)	1.85 (119.01)	40.88 (1.03)	0.73 (0.31)	0.62 (0.46)	1.27	1673.53
Ajedrez	C	30.0	0.69 (0.07)	1.86 (112.27)	42.15 (1.05)	0.66 (0.29)	0.52 (0.41)	1.34	1672.50

**Tabla A-12.:** Promedio y desviación estándar (en paréntesis) de  $ABC$ ,  $d(\beta, \hat{\beta})$ ,  $d_2(\beta_v, \hat{\beta}_v)$ ,  $d(\theta, \hat{\theta})$ ,  $d_2(\theta, \hat{\theta})$ ,  $\hat{\theta}$  y tiempo total de simulación de 200 replicaciones. Modelo de respuesta binaria,  $n = 220$ , bajo niveles de contaminación  $\gamma_c$  y contaminación de cambio de forma - Tipo 1 en parámetro imagen.

$\beta_0$	$\gamma_c(\%)$	<i>RMSPE</i>	$d(\beta, \hat{\beta})$	$d_2(\beta_v, \hat{\beta}_v)$	$d(\theta, \hat{\theta})$	$d_2(\theta, \hat{\theta})$	$\hat{\theta}$	Tiempo (Seg)
T	5	3.14 (0.42)	1.22 (67.90)	30.27 (0.83)	0.31 (0.22)	0.14 (0.18)	1.79	1226.59
T	15	3.21 (0.42)	1.27 (82.38)	30.43 (0.87)	0.34 (0.24)	0.17 (0.22)	1.75	1228.30
T	30	3.21 (0.44)	1.32 (84.61)	30.92 (0.89)	0.34 (0.24)	0.17 (0.21)	1.77	1228.94
Triangulo	5	2.30 (0.31)	0.93 (48.30)	20.84 (0.60)	0.24 (0.16)	0.08 (0.10)	1.83	1230.89
Triangulo	15	2.29 (0.31)	0.96 (49.46)	21.02 (0.69)	0.24 (0.18)	0.09 (0.11)	1.84	1234.28
Triangulo	30	2.34 (0.28)	0.99 (53.72)	21.27 (0.63)	0.26 (0.20)	0.11 (0.14)	1.80	1234.00
Cuadrado	5	2.93 (0.39)	1.06 (58.88)	26.74 (0.69)	0.31 (0.22)	0.14 (0.18)	1.76	1234.25
Cuadrado	15	2.87 (0.38)	1.12 (73.87)	26.93 (0.77)	0.32 (0.22)	0.15 (0.18)	1.77	1233.47
Cuadrado	30	2.93 (0.39)	1.16 (68.90)	27.18 (0.76)	0.32 (0.23)	0.15 (0.19)	1.76	1238.06
Ajedrez	5	3.85 (0.50)	1.36 (76.31)	37.39 (0.93)	0.36 (0.25)	0.20 (0.23)	1.76	1235.78
Ajedrez	15	3.87 (0.49)	1.36 (71.76)	37.36 (1.06)	0.39 (0.27)	0.22 (0.28)	1.75	1230.08
Ajedrez	30	3.85 (0.53)	1.38 (78.81)	37.34 (0.99)	0.37 (0.30)	0.23 (0.33)	1.72	1237.67

**Tabla A-13.:** Promedio y desviación estándar (en paréntesis) de *RMSPE*,  $d(\beta, \hat{\beta})$ ,  $d_2(\beta_v, \hat{\beta}_v)$ ,  $d(\theta, \hat{\theta})$ ,  $d_2(\theta, \hat{\theta})$ ,  $\hat{\theta}$  y tiempo total de simulación de 200 replicas. Modelo de respuesta continua,  $n = 100$ , bajo niveles de contaminación  $\gamma_c$  y contaminación de cambio de forma - Tipo 2 en parámetro imagen (zoom 50% de parámetro imagen original).

$\beta_0$	$\gamma_c(\%)$	<i>RMSPE</i>	$d(\beta, \hat{\beta})$	$d_2(\beta_v, \hat{\beta}_v)$	$d(\theta, \hat{\theta})$	$d_2(\theta, \hat{\theta})$	$\hat{\theta}$	Tiempo (Seg)
T	5	2.94 (0.30)	1.29 (51.77)	27.60 (0.83)	0.22 (0.17)	0.08 (0.10)	1.86	1957.19
T	15	3.02 (0.32)	1.35 (61.15)	28.27 (0.87)	0.23 (0.18)	0.08 (0.13)	1.86	1971.42
T	30	3.07 (0.32)	1.43 (70.05)	28.95 (0.87)	0.22 (0.17)	0.08 (0.12)	1.87	1979.19
Triangulo	5	2.13 (0.23)	0.98 (41.78)	18.84 (0.60)	0.16 (0.12)	0.04 (0.05)	1.90	1986.23
Triangulo	15	2.15 (0.23)	1.02 (44.66)	19.23 (0.62)	0.17 (0.12)	0.04 (0.05)	1.88	1982.53
Triangulo	30	2.21 (0.24)	1.07 (50.22)	19.78 (0.69)	0.18 (0.14)	0.05 (0.07)	1.88	1988.33
Cuadrado	5	2.61 (0.27)	1.13 (48.40)	24.11 (0.76)	0.18 (0.14)	0.05 (0.07)	1.87	1978.83
Cuadrado	15	2.63 (0.28)	1.18 (53.58)	24.67 (0.82)	0.20 (0.15)	0.06 (0.09)	1.88	1982.78
Cuadrado	30	2.72 (0.28)	1.25 (58.16)	25.19 (0.77)	0.23 (0.16)	0.08 (0.09)	1.86	1975.47
Ajedrez	5	3.59 (0.36)	1.43 (61.35)	34.09 (1.12)	0.24 (0.17)	0.09 (0.11)	1.86	1978.03
Ajedrez	15	3.55 (0.37)	1.44 (58.93)	34.15 (1.14)	0.25 (0.18)	0.09 (0.13)	1.84	1980.77
Ajedrez	30	3.54 (0.38)	1.42 (58.84)	34.19 (1.16)	0.24 (0.19)	0.09 (0.13)	1.85	1968.45

**Tabla A-14.:** Promedio y desviación estándar (en paréntesis) de *RMSPE*,  $d(\beta, \hat{\beta})$ ,  $d_2(\beta_v, \hat{\beta}_v)$ ,  $d(\theta, \hat{\theta})$ ,  $d_2(\theta, \hat{\theta})$ ,  $\hat{\theta}$  y tiempo total de simulación de 200 replicas. Modelo de respuesta continua,  $n = 170$ , bajo niveles de contaminación  $\gamma_c$  y contaminación de cambio de forma - Tipo 2 en parámetro imagen (zoom 50% de parámetro imagen original).

$\beta_0$	$\gamma_c(\%)$	<i>RMSPE</i>	$d(\beta, \hat{\beta})$	$d_2(\beta_v, \hat{\beta}_v)$	$d(\theta, \hat{\theta})$	$d_2(\theta, \hat{\theta})$	$\hat{\theta}$	Tiempo (Seg)
T	5	2.81 (0.26)	1.31 (52.65)	26.21 (0.84)	0.19 (0.14)	0.06 (0.07)	1.88	2517.00
T	15	2.85 (0.28)	1.39 (59.37)	26.94 (0.82)	0.20 (0.14)	0.06 (0.08)	1.86	2526.58
T	30	2.96 (0.27)	1.46 (66.81)	27.81 (0.90)	0.21 (0.15)	0.07 (0.09)	1.87	2528.22
Triangulo	5	2.01 (0.19)	1.00 (42.27)	17.75 (0.62)	0.12 (0.10)	0.02 (0.04)	1.93	2538.31
Triangulo	15	2.08 (0.18)	1.05 (47.10)	18.31 (0.61)	0.14 (0.10)	0.03 (0.04)	1.92	2568.20
Triangulo	30	2.16 (0.20)	1.10 (42.11)	19.01 (0.67)	0.14 (0.10)	0.03 (0.04)	1.91	2530.62
Cuadrado	5	2.47 (0.21)	1.15 (46.39)	22.59 (0.73)	0.15 (0.12)	0.04 (0.05)	1.91	2523.69
Cuadrado	15	2.49 (0.21)	1.22 (52.48)	23.15 (0.76)	0.15 (0.11)	0.04 (0.05)	1.91	2530.45
Cuadrado	30	2.62 (0.24)	1.29 (56.34)	24.14 (0.81)	0.18 (0.12)	0.05 (0.06)	1.88	2540.03
Ajedrez	5	3.39 (0.32)	1.45 (53.33)	32.23 (1.10)	0.19 (0.14)	0.05 (0.07)	1.90	2535.80
Ajedrez	15	3.37 (0.29)	1.45 (52.90)	32.14 (1.15)	0.20 (0.15)	0.06 (0.09)	1.89	2538.00
Ajedrez	30	3.40 (0.31)	1.46 (46.89)	32.05 (1.10)	0.21 (0.16)	0.07 (0.10)	1.87	2544.67

**Tabla A-15.:** Promedio y desviación estándar (en paréntesis) de *RMSPE*,  $d(\beta, \hat{\beta})$ ,  $d_2(\beta_v, \hat{\beta}_v)$ ,  $d(\theta, \hat{\theta})$ ,  $d_2(\theta, \hat{\theta})$ ,  $\hat{\theta}$  y tiempo total de simulación de 200 replicas. Modelo de respuesta continua,  $n = 220$ , bajo niveles de contaminación  $\gamma_c$  y contaminación de cambio de forma - Tipo 2 en parámetro imagen (zoom 50% de parámetro imagen original).

$\beta_0$	$\gamma_c$ (%)	$\gamma_b$	$RMSPE$	$v_c$				$f_c$				$RMSPE$ (sin influyentes)			
				$DO$	$DO_b(0.995)$	$DO_b(0.95)$	$MS$	$DO$	$DO_b(0.995)$	$DO_b(0.95)$	$MS$	$DO$	$DO_b(0.995)$	$DO_b(0.95)$	$MS$
T	0	1	2.32	-	-	-	-	0.00	0.02	0.05	0.10	2.32	2.45	2.54	2.63
T	0	1	(0.47)	-	-	-	-	(0.00)	(0.01)	(0.01)	(0.06)	(0.47)	(0.48)	(0.50)	(0.50)
T	5	2	2.27	0.04	0.21	0.34	0.40	0.00	0.01	0.04	0.10	2.28	2.39	2.48	2.60
T	5	2	(0.41)	(0.15)	(0.26)	(0.32)	(0.35)	(0.01)	(0.01)	(0.02)	(0.05)	(0.43)	(0.43)	(0.45)	(0.48)
T	5	5	2.55	0.37	0.51	0.70	0.72	0.00	0.00	0.02	0.09	2.37	2.38	2.43	2.56
T	5	5	(0.57)	(0.37)	(0.22)	(0.29)	(0.30)	(0.01)	(0.01)	(0.02)	(0.06)	(0.44)	(0.46)	(0.47)	(0.50)
T	15	2	2.36	0.04	0.14	0.27	0.42	0.00	0.00	0.02	0.08	2.40	2.46	2.52	2.68
T	15	2	(0.46)	(0.09)	(0.07)	(0.11)	(0.19)	(0.00)	(0.01)	(0.02)	(0.06)	(0.48)	(0.51)	(0.50)	(0.56)
T	15	5	3.08	0.31	0.17	0.41	0.72	0.00	0.00	0.00	0.07	2.65	2.71	2.51	2.55
T	15	5	(0.70)	(0.24)	(0.06)	(0.08)	(0.16)	(0.01)	(0.00)	(0.00)	(0.05)	(0.55)	(0.58)	(0.51)	(0.52)
T	30	2	2.28	0.02	0.07	0.17	0.36	0.00	0.00	0.00	0.05	2.31	2.38	2.44	2.60
T	30	2	(0.45)	(0.05)	(0.02)	(0.05)	(0.13)	(0.00)	(0.01)	(0.01)	(0.05)	(0.45)	(0.50)	(0.53)	(0.56)
T	30	5	3.97	0.12	0.08	0.19	0.58	0.00	0.00	0.00	0.03	3.58	3.45	3.16	2.70
T	30	5	(1.09)	(0.13)	(0.03)	(0.04)	(0.12)	(0.00)	(0.00)	(0.00)	(0.04)	(1.14)	(0.76)	(0.70)	(0.58)
Triangulo	0	1	1.86	-	-	-	-	0.00	0.02	0.05	0.12	1.87	1.90	1.94	2.01
Triangulo	0	1	(0.38)	-	-	-	-	(0.00)	(0.01)	(0.01)	(0.07)	(0.37)	(0.36)	(0.36)	(0.38)
Triangulo	5	2	1.86	0.02	0.21	0.30	0.40	0.00	0.02	0.04	0.11	1.85	1.91	1.94	2.05
Triangulo	5	2	(0.32)	(0.12)	(0.26)	(0.32)	(0.36)	(0.00)	(0.01)	(0.02)	(0.07)	(0.33)	(0.32)	(0.33)	(0.37)
Triangulo	5	5	1.92	0.34	0.49	0.70	0.74	0.00	0.00	0.02	0.10	1.85	1.83	1.85	1.96
Triangulo	5	5	(0.35)	(0.36)	(0.26)	(0.34)	(0.33)	(0.02)	(0.01)	(0.02)	(0.07)	(0.34)	(0.33)	(0.33)	(0.34)
Triangulo	15	2	1.82	0.02	0.12	0.23	0.40	0.00	0.01	0.02	0.09	1.85	1.88	1.93	2.02
Triangulo	15	2	(0.36)	(0.06)	(0.06)	(0.11)	(0.18)	(0.00)	(0.01)	(0.02)	(0.06)	(0.36)	(0.33)	(0.33)	(0.32)
Triangulo	15	5	2.36	0.28	0.17	0.37	0.70	0.00	0.00	0.00	0.07	2.08	2.13	2.00	1.99
Triangulo	15	5	(0.49)	(0.22)	(0.05)	(0.08)	(0.18)	(0.01)	(0.00)	(0.00)	(0.05)	(0.39)	(0.39)	(0.36)	(0.33)
Triangulo	30	2	1.85	0.01	0.07	0.16	0.35	0.00	0.00	0.01	0.07	1.85	1.89	1.92	2.02
Triangulo	30	2	(0.37)	(0.04)	(0.03)	(0.05)	(0.13)	(0.00)	(0.01)	(0.02)	(0.06)	(0.38)	(0.38)	(0.39)	(0.41)
Triangulo	30	5	2.82	0.09	0.08	0.19	0.58	0.00	0.00	0.00	0.04	2.57	2.54	2.27	2.01
Triangulo	30	5	(0.68)	(0.12)	(0.02)	(0.04)	(0.12)	(0.00)	(0.00)	(0.00)	(0.05)	(0.68)	(0.52)	(0.42)	(0.39)
Cuadrado	0	1	2.02	-	-	-	-	0.00	0.02	0.05	0.11	2.03	2.14	2.22	2.37
Cuadrado	0	1	(0.39)	-	-	-	-	(0.00)	(0.01)	(0.01)	(0.07)	(0.39)	(0.41)	(0.40)	(0.43)
Cuadrado	5	2	1.93	0.06	0.20	0.34	0.44	0.00	0.02	0.04	0.10	1.93	2.04	2.13	2.28
Cuadrado	5	2	(0.39)	(0.17)	(0.25)	(0.29)	(0.34)	(0.01)	(0.01)	(0.02)	(0.06)	(0.39)	(0.40)	(0.40)	(0.47)
Cuadrado	5	5	2.12	0.45	0.50	0.73	0.81	0.00	0.00	0.02	0.09	2.03	2.02	2.06	2.21
Cuadrado	5	5	(0.44)	(0.41)	(0.22)	(0.28)	(0.26)	(0.01)	(0.01)	(0.01)	(0.06)	(0.39)	(0.41)	(0.43)	(0.45)
Cuadrado	15	2	1.95	0.05	0.14	0.26	0.42	0.00	0.00	0.02	0.08	1.99	2.03	2.11	2.25
Cuadrado	15	2	(0.43)	(0.11)	(0.06)	(0.11)	(0.19)	(0.00)	(0.01)	(0.02)	(0.06)	(0.43)	(0.47)	(0.49)	(0.50)
Cuadrado	15	5	2.75	0.35	0.16	0.39	0.71	0.00	0.00	0.00	0.06	2.18	2.30	2.14	2.10
Cuadrado	15	5	(0.76)	(0.20)	(0.05)	(0.09)	(0.17)	(0.01)	(0.00)	(0.01)	(0.05)	(0.45)	(0.46)	(0.40)	(0.39)
Cuadrado	30	2	1.97	0.03	0.07	0.17	0.37	0.00	0.00	0.00	0.05	1.99	2.05	2.12	2.26
Cuadrado	30	2	(0.42)	(0.06)	(0.03)	(0.05)	(0.14)	(0.00)	(0.00)	(0.01)	(0.04)	(0.42)	(0.43)	(0.44)	(0.52)
Cuadrado	30	5	3.67	0.12	0.08	0.18	0.57	0.00	0.00	0.00	0.04	3.20	3.24	2.85	2.36
Cuadrado	30	5	(1.12)	(0.14)	(0.03)	(0.04)	(0.15)	(0.00)	(0.00)	(0.00)	(0.05)	(1.00)	(0.83)	(0.73)	(0.53)
Ajedrez	0	1	2.59	-	-	-	-	0.00	0.02	0.05	0.10	2.59	2.73	2.86	3.03
Ajedrez	0	1	(0.58)	-	-	-	-	(0.00)	(0.01)	(0.01)	(0.06)	(0.58)	(0.59)	(0.62)	(0.66)
Ajedrez	5	2	2.64	0.07	0.27	0.35	0.44	0.00	0.01	0.04	0.09	2.65	2.74	2.83	3.01
Ajedrez	5	2	(0.50)	(0.19)	(0.27)	(0.31)	(0.33)	(0.01)	(0.01)	(0.02)	(0.06)	(0.52)	(0.53)	(0.57)	(0.62)
Ajedrez	5	5	2.97	0.42	0.48	0.72	0.80	0.00	0.00	0.02	0.08	2.84	2.84	2.89	3.09
Ajedrez	5	5	(0.74)	(0.38)	(0.18)	(0.30)	(0.30)	(0.01)	(0.01)	(0.02)	(0.06)	(0.63)	(0.60)	(0.66)	(0.67)
Ajedrez	15	2	2.61	0.08	0.15	0.28	0.45	0.00	0.00	0.01	0.07	2.65	2.75	2.83	3.08
Ajedrez	15	2	(0.48)	(0.15)	(0.07)	(0.12)	(0.19)	(0.00)	(0.01)	(0.02)	(0.05)	(0.51)	(0.51)	(0.50)	(0.59)
Ajedrez	15	5	3.78	0.35	0.17	0.39	0.72	0.00	0.00	0.00	0.07	3.00	3.18	2.90	2.82
Ajedrez	15	5	(0.94)	(0.21)	(0.06)	(0.10)	(0.16)	(0.01)	(0.00)	(0.00)	(0.05)	(0.75)	(0.73)	(0.56)	(0.62)
Ajedrez	30	2	2.62	0.03	0.08	0.16	0.37	0.00	0.00	0.00	0.06	2.69	2.72	2.80	3.11
Ajedrez	30	2	(0.47)	(0.07)	(0.03)	(0.05)	(0.15)	(0.00)	(0.00)	(0.01)	(0.04)	(0.48)	(0.51)	(0.57)	(0.64)
Ajedrez	30	5	4.98	0.10	0.08	0.18	0.55	0.00	0.00	0.00	0.03	4.36	4.21	3.74	3.04
Ajedrez	30	5	(1.42)	(0.13)	(0.03)	(0.04)	(0.14)	(0.00)	(0.00)	(0.00)	(0.04)	(1.37)	(1.12)	(0.94)	(0.68)

**Tabla A-16.:** Promedio y desviación estándar (en paréntesis) de  $RMSPE$ , tasa de correcta detección ( $v_c$ ), tasa de falsa detección ( $f_c$ ) de cada método de identificación de influyentes, en 100 replicaciones. Modelo de respuesta continua,  $n = 50$ , bajo niveles de contaminación  $\gamma_c$  y factor de contaminación de magnitud  $\gamma_b$ .



$\beta_0$	$\gamma_c(\%)$	$\gamma_b$	$RMSPE$	$v_c$				$f_c$				$RMSPE$ (sin influentes)			
				$DO$	$DO_b(0.995)$	$DO_b(0.95)$	$MS$	$DO$	$DO_b(0.995)$	$DO_b(0.95)$	$MS$	$DO$	$DO_b(0.995)$	$DO_b(0.95)$	$MS$
T	0	1	2.11	-	-	-	-	0.00	0.02	0.05	0.12	2.12	2.20	2.33	2.49
T	0	1	(0.38)	-	-	-	-	(0.00)	(0.00)	(0.01)	(0.06)	(0.38)	(0.39)	(0.41)	(0.44)
T	5	2	2.14	0.12	0.24	0.43	0.53	0.00	0.01	0.04	0.11	2.18	2.24	2.35	2.51
T	5	2	(0.36)	(0.23)	(0.18)	(0.30)	(0.31)	(0.00)	(0.01)	(0.02)	(0.05)	(0.37)	(0.39)	(0.42)	(0.43)
T	5	5	2.46	0.50	0.38	0.69	0.76	0.00	0.00	0.02	0.10	2.20	2.23	2.20	2.37
T	5	5	(0.51)	(0.29)	(0.14)	(0.25)	(0.24)	(0.01)	(0.00)	(0.01)	(0.06)	(0.38)	(0.41)	(0.37)	(0.43)
T	15	2	2.07	0.07	0.12	0.28	0.44	0.00	0.00	0.01	0.07	2.11	2.13	2.20	2.35
T	15	2	(0.32)	(0.11)	(0.04)	(0.09)	(0.17)	(0.00)	(0.00)	(0.01)	(0.05)	(0.33)	(0.33)	(0.36)	(0.39)
T	15	5	3.19	0.34	0.12	0.37	0.72	0.00	0.00	0.00	0.07	2.43	2.73	2.39	2.31
T	15	5	(0.68)	(0.19)	(0.04)	(0.07)	(0.15)	(0.01)	(0.00)	(0.00)	(0.05)	(0.47)	(0.54)	(0.42)	(0.44)
T	30	2	2.11	0.03	0.05	0.17	0.39	0.00	0.00	0.00	0.06	2.15	2.16	2.21	2.34
T	30	2	(0.33)	(0.06)	(0.02)	(0.04)	(0.11)	(0.01)	(0.00)	(0.01)	(0.05)	(0.33)	(0.34)	(0.36)	(0.38)
T	30	5	4.31	0.09	0.05	0.18	0.54	0.00	0.00	0.00	0.04	3.58	3.72	3.11	2.47
T	30	5	(1.04)	(0.11)	(0.02)	(0.03)	(0.12)	(0.00)	(0.00)	(0.00)	(0.04)	(0.92)	(0.85)	(0.69)	(0.42)
Triangulo	0	1	1.72	-	-	-	-	0.00	0.02	0.05	0.13	1.73	1.76	1.82	1.91
Triangulo	0	1	(0.29)	-	-	-	-	(0.00)	(0.01)	(0.01)	(0.05)	(0.29)	(0.30)	(0.32)	(0.35)
Triangulo	5	2	1.70	0.04	0.18	0.34	0.45	0.00	0.01	0.04	0.12	1.71	1.74	1.81	1.88
Triangulo	5	2	(0.27)	(0.12)	(0.19)	(0.24)	(0.30)	(0.01)	(0.01)	(0.01)	(0.05)	(0.28)	(0.26)	(0.28)	(0.28)
Triangulo	5	5	1.89	0.40	0.38	0.68	0.77	0.00	0.00	0.02	0.10	1.74	1.76	1.75	1.84
Triangulo	5	5	(0.30)	(0.29)	(0.13)	(0.24)	(0.25)	(0.01)	(0.00)	(0.02)	(0.05)	(0.30)	(0.30)	(0.31)	(0.33)
Triangulo	15	2	1.67	0.03	0.10	0.23	0.39	0.00	0.00	0.02	0.09	1.69	1.72	1.80	1.88
Triangulo	15	2	(0.28)	(0.07)	(0.04)	(0.09)	(0.16)	(0.00)	(0.01)	(0.02)	(0.05)	(0.28)	(0.29)	(0.31)	(0.33)
Triangulo	15	5	2.40	0.28	0.12	0.37	0.68	0.00	0.00	0.00	0.07	2.00	2.13	1.93	1.83
Triangulo	15	5	(0.47)	(0.17)	(0.04)	(0.07)	(0.14)	(0.01)	(0.00)	(0.00)	(0.05)	(0.36)	(0.37)	(0.32)	(0.30)
Triangulo	30	2	1.72	0.02	0.06	0.16	0.36	0.00	0.00	0.01	0.07	1.73	1.76	1.81	1.93
Triangulo	30	2	(0.27)	(0.04)	(0.02)	(0.03)	(0.11)	(0.00)	(0.00)	(0.01)	(0.05)	(0.27)	(0.27)	(0.29)	(0.33)
Triangulo	30	5	3.06	0.10	0.05	0.18	0.55	0.00	0.00	0.00	0.05	2.63	2.68	2.32	1.98
Triangulo	30	5	(0.71)	(0.10)	(0.02)	(0.03)	(0.12)	(0.00)	(0.00)	(0.00)	(0.04)	(0.60)	(0.52)	(0.41)	(0.30)
Cuadrado	0	1	1.75	-	-	-	-	0.00	0.02	0.05	0.12	1.77	1.86	1.96	2.10
Cuadrado	0	1	(0.30)	-	-	-	-	(0.01)	(0.00)	(0.01)	(0.05)	(0.32)	(0.32)	(0.34)	(0.38)
Cuadrado	5	2	1.77	0.08	0.22	0.43	0.52	0.00	0.01	0.03	0.10	1.78	1.84	1.93	2.07
Cuadrado	5	2	(0.30)	(0.18)	(0.17)	(0.28)	(0.28)	(0.01)	(0.01)	(0.01)	(0.04)	(0.32)	(0.33)	(0.35)	(0.36)
Cuadrado	5	5	2.12	0.51	0.38	0.74	0.80	0.00	0.00	0.02	0.09	1.88	1.91	1.87	2.02
Cuadrado	5	5	(0.45)	(0.33)	(0.14)	(0.24)	(0.25)	(0.01)	(0.00)	(0.01)	(0.06)	(0.30)	(0.30)	(0.32)	(0.35)
Cuadrado	15	2	1.77	0.08	0.10	0.27	0.47	0.00	0.00	0.01	0.08	1.80	1.82	1.90	2.05
Cuadrado	15	2	(0.30)	(0.13)	(0.04)	(0.09)	(0.16)	(0.01)	(0.01)	(0.01)	(0.04)	(0.29)	(0.32)	(0.32)	(0.38)
Cuadrado	15	5	2.87	0.36	0.12	0.37	0.72	0.00	0.00	0.00	0.06	2.17	2.49	2.17	1.99
Cuadrado	15	5	(0.66)	(0.18)	(0.04)	(0.07)	(0.14)	(0.01)	(0.00)	(0.00)	(0.05)	(0.40)	(0.58)	(0.40)	(0.38)
Cuadrado	30	2	1.85	0.03	0.05	0.16	0.38	0.00	0.00	0.00	0.06	1.86	1.88	1.93	2.10
Cuadrado	30	2	(0.31)	(0.06)	(0.02)	(0.03)	(0.11)	(0.00)	(0.00)	(0.01)	(0.05)	(0.33)	(0.37)	(0.38)	(0.43)
Cuadrado	30	5	3.64	0.13	0.06	0.18	0.55	0.00	0.00	0.00	0.04	3.01	3.19	2.68	2.16
Cuadrado	30	5	(0.98)	(0.11)	(0.02)	(0.03)	(0.12)	(0.00)	(0.00)	(0.00)	(0.04)	(0.95)	(0.80)	(0.55)	(0.38)
Ajedrez	0	1	2.43	-	-	-	-	0.00	0.02	0.05	0.12	2.44	2.51	2.67	2.85
Ajedrez	0	1	(0.46)	-	-	-	-	(0.01)	(0.01)	(0.01)	(0.06)	(0.45)	(0.46)	(0.48)	(0.55)
Ajedrez	5	2	2.42	0.09	0.26	0.43	0.53	0.00	0.01	0.03	0.10	2.43	2.48	2.58	2.81
Ajedrez	5	2	(0.38)	(0.17)	(0.19)	(0.28)	(0.28)	(0.00)	(0.01)	(0.02)	(0.06)	(0.36)	(0.38)	(0.39)	(0.50)
Ajedrez	5	5	2.84	0.56	0.37	0.77	0.84	0.00	0.00	0.01	0.08	2.44	2.51	2.44	2.58
Ajedrez	5	5	(0.67)	(0.33)	(0.13)	(0.20)	(0.19)	(0.01)	(0.00)	(0.01)	(0.05)	(0.41)	(0.48)	(0.42)	(0.43)
Ajedrez	15	2	2.40	0.09	0.11	0.31	0.50	0.00	0.00	0.01	0.07	2.40	2.43	2.50	2.69
Ajedrez	15	2	(0.41)	(0.12)	(0.05)	(0.08)	(0.16)	(0.01)	(0.00)	(0.01)	(0.05)	(0.41)	(0.43)	(0.45)	(0.53)
Ajedrez	15	5	3.86	0.34	0.12	0.39	0.70	0.00	0.00	0.00	0.06	2.87	3.25	2.70	2.67
Ajedrez	15	5	(1.05)	(0.17)	(0.04)	(0.07)	(0.13)	(0.01)	(0.00)	(0.00)	(0.04)	(0.67)	(0.74)	(0.48)	(0.49)
Ajedrez	30	2	2.46	0.06	0.05	0.17	0.41	0.00	0.00	0.00	0.06	2.48	2.46	2.52	2.71
Ajedrez	30	2	(0.39)	(0.09)	(0.02)	(0.04)	(0.12)	(0.00)	(0.00)	(0.01)	(0.05)	(0.41)	(0.41)	(0.45)	(0.51)
Ajedrez	30	5	5.31	0.09	0.06	0.18	0.56	0.00	0.00	0.00	0.05	4.45	4.55	3.81	2.90
Ajedrez	30	5	(1.19)	(0.11)	(0.02)	(0.03)	(0.12)	(0.00)	(0.00)	(0.00)	(0.04)	(1.24)	(1.02)	(0.86)	(0.57)

**Tabla A-17.:** Promedio y desviación estándar (en paréntesis) de  $RMSPE$ , tasa de correcta detección ( $v_c$ ), tasa de falsa detección ( $f_c$ ) de cada método de identificación de influentes, en 100 repeticiones. Modelo de respuesta continua,  $n = 70$ , bajo niveles de contaminación  $\gamma_c$  y factor de contaminación de magnitud  $\gamma_b$ .

$\beta_0$	$\gamma_c$ (%)	$\gamma_b$	ABC	$v_c$				$f_c$				ABC (sin influentes)			
				DO	DO <sub>b</sub> (0.995)	DO <sub>b</sub> (0.95)	MS	DO	DO <sub>b</sub> (0.995)	DO <sub>b</sub> (0.95)	MS	DO	DO <sub>b</sub> (0.995)	DO <sub>b</sub> (0.95)	MS
T	0	1	0.78	-	-	-	-	0.00	0.02	0.05	0.01	0.78	0.78	0.79	0.78
T	0	1	(0.14)	-	-	-	-	(0.00)	(0.01)	(0.01)	(0.01)	(0.14)	(0.14)	(0.14)	(0.14)
T	5	2	0.81	0.00	0.03	0.04	0.01	0.00	0.02	0.05	0.01	0.81	0.80	0.81	0.80
T	5	2	(0.12)	(0.00)	(0.12)	(0.14)	(0.07)	(0.00)	(0.01)	(0.01)	(0.02)	(0.12)	(0.12)	(0.12)	(0.12)
T	5	5	0.78	0.00	0.02	0.04	0.00	0.00	0.02	0.06	0.01	0.78	0.78	0.79	0.78
T	5	5	(0.13)	(0.00)	(0.10)	(0.14)	(0.05)	(0.00)	(0.01)	(0.02)	(0.02)	(0.13)	(0.12)	(0.12)	(0.13)
T	15	2	0.79	0.00	0.02	0.04	0.00	0.00	0.02	0.05	0.01	0.79	0.79	0.79	0.79
T	15	2	(0.13)	(0.00)	(0.05)	(0.06)	(0.02)	(0.00)	(0.01)	(0.02)	(0.02)	(0.13)	(0.13)	(0.13)	(0.13)
T	15	5	0.77	0.00	0.01	0.05	0.01	0.00	0.02	0.06	0.01	0.77	0.77	0.77	0.77
T	15	5	(0.14)	(0.00)	(0.04)	(0.08)	(0.03)	(0.00)	(0.01)	(0.02)	(0.02)	(0.14)	(0.15)	(0.14)	(0.14)
T	30	2	0.81	0.00	0.02	0.04	0.01	0.00	0.02	0.06	0.01	0.81	0.80	0.80	0.81
T	30	2	(0.11)	(0.00)	(0.03)	(0.05)	(0.03)	(0.00)	(0.01)	(0.03)	(0.02)	(0.11)	(0.11)	(0.11)	(0.11)
T	30	5	0.78	0.00	0.02	0.05	0.01	0.00	0.02	0.06	0.01	0.78	0.78	0.78	0.78
T	30	5	(0.11)	(0.00)	(0.03)	(0.05)	(0.02)	(0.00)	(0.02)	(0.02)	(0.02)	(0.11)	(0.10)	(0.11)	(0.10)
Triangulo	0	1	0.71	-	-	-	-	0.00	0.02	0.06	0.01	0.71	0.73	0.71	0.72
Triangulo	0	1	(0.16)	-	-	-	-	(0.00)	(0.01)	(0.01)	(0.01)	(0.16)	(0.15)	(0.15)	(0.16)
Triangulo	5	2	0.70	0.00	0.04	0.07	0.02	0.00	0.02	0.05	0.01	0.70	0.70	0.72	0.70
Triangulo	5	2	(0.15)	(0.00)	(0.14)	(0.18)	(0.10)	(0.00)	(0.01)	(0.01)	(0.01)	(0.15)	(0.14)	(0.15)	(0.16)
Triangulo	5	5	0.73	0.00	0.04	0.05	0.00	0.00	0.02	0.06	0.01	0.73	0.73	0.72	0.73
Triangulo	5	5	(0.13)	(0.00)	(0.13)	(0.15)	(0.05)	(0.00)	(0.01)	(0.01)	(0.01)	(0.13)	(0.14)	(0.14)	(0.13)
Triangulo	15	2	0.72	0.00	0.02	0.06	0.01	0.00	0.02	0.06	0.01	0.72	0.71	0.72	0.71
Triangulo	15	2	(0.14)	(0.00)	(0.05)	(0.09)	(0.04)	(0.00)	(0.01)	(0.02)	(0.02)	(0.14)	(0.15)	(0.14)	(0.14)
Triangulo	15	5	0.73	0.00	0.02	0.05	0.00	0.00	0.02	0.05	0.01	0.73	0.72	0.74	0.73
Triangulo	15	5	(0.14)	(0.01)	(0.05)	(0.07)	(0.01)	(0.00)	(0.01)	(0.02)	(0.02)	(0.14)	(0.15)	(0.14)	(0.14)
Triangulo	30	2	0.74	0.00	0.02	0.06	0.01	0.00	0.02	0.05	0.01	0.74	0.74	0.74	0.74
Triangulo	30	2	(0.13)	(0.01)	(0.03)	(0.05)	(0.02)	(0.00)	(0.02)	(0.02)	(0.02)	(0.13)	(0.13)	(0.14)	(0.13)
Triangulo	30	5	0.73	0.00	0.02	0.05	0.01	0.00	0.03	0.06	0.01	0.73	0.74	0.72	0.73
Triangulo	30	5	(0.13)	(0.00)	(0.03)	(0.05)	(0.03)	(0.00)	(0.02)	(0.03)	(0.02)	(0.13)	(0.14)	(0.14)	(0.13)
Cuadrado	0	1	0.79	-	-	-	-	0.00	0.02	0.05	0.01	0.79	0.79	0.79	0.78
Cuadrado	0	1	(0.14)	-	-	-	-	(0.00)	(0.01)	(0.01)	(0.01)	(0.14)	(0.14)	(0.14)	(0.13)
Cuadrado	5	2	0.79	0.00	0.01	0.04	0.00	0.00	0.02	0.05	0.01	0.79	0.79	0.79	0.79
Cuadrado	5	2	(0.12)	(0.00)	(0.09)	(0.13)	(0.05)	(0.00)	(0.01)	(0.01)	(0.02)	(0.12)	(0.11)	(0.10)	(0.12)
Cuadrado	5	5	0.79	0.00	0.01	0.03	0.00	0.00	0.02	0.06	0.01	0.79	0.79	0.79	0.79
Cuadrado	5	5	(0.13)	(0.00)	(0.09)	(0.12)	(0.05)	(0.00)	(0.01)	(0.02)	(0.02)	(0.13)	(0.13)	(0.14)	(0.13)
Cuadrado	15	2	0.78	0.00	0.01	0.05	0.01	0.00	0.02	0.06	0.01	0.78	0.78	0.79	0.78
Cuadrado	15	2	(0.13)	(0.00)	(0.04)	(0.08)	(0.03)	(0.00)	(0.01)	(0.02)	(0.02)	(0.13)	(0.13)	(0.14)	(0.13)
Cuadrado	15	5	0.78	0.00	0.02	0.03	0.01	0.00	0.02	0.06	0.01	0.78	0.78	0.78	0.77
Cuadrado	15	5	(0.14)	(0.00)	(0.05)	(0.06)	(0.04)	(0.00)	(0.01)	(0.02)	(0.02)	(0.14)	(0.13)	(0.13)	(0.14)
Cuadrado	30	2	0.80	0.00	0.01	0.04	0.01	0.00	0.03	0.06	0.01	0.80	0.79	0.79	0.79
Cuadrado	30	2	(0.14)	(0.00)	(0.03)	(0.04)	(0.02)	(0.00)	(0.01)	(0.02)	(0.03)	(0.14)	(0.14)	(0.14)	(0.14)
Cuadrado	30	5	0.79	0.00	0.01	0.04	0.01	0.00	0.03	0.06	0.01	0.79	0.80	0.79	0.79
Cuadrado	30	5	(0.14)	(0.00)	(0.03)	(0.04)	(0.02)	(0.00)	(0.01)	(0.02)	(0.02)	(0.14)	(0.13)	(0.13)	(0.14)
Ajedrez	0	1	0.82	-	-	-	-	0.00	0.02	0.05	0.01	0.82	0.81	0.81	0.82
Ajedrez	0	1	(0.13)	-	-	-	-	(0.00)	(0.01)	(0.01)	(0.02)	(0.13)	(0.13)	(0.14)	(0.13)
Ajedrez	5	2	0.84	0.00	0.01	0.03	0.00	0.00	0.02	0.06	0.01	0.84	0.84	0.84	0.84
Ajedrez	5	2	(0.12)	(0.00)	(0.09)	(0.12)	(0.00)	(0.00)	(0.01)	(0.01)	(0.02)	(0.12)	(0.12)	(0.12)	(0.12)
Ajedrez	5	5	0.82	0.00	0.01	0.03	0.00	0.00	0.02	0.05	0.01	0.82	0.83	0.83	0.81
Ajedrez	5	5	(0.13)	(0.00)	(0.07)	(0.14)	(0.05)	(0.00)	(0.01)	(0.01)	(0.02)	(0.13)	(0.12)	(0.12)	(0.13)
Ajedrez	15	2	0.83	0.00	0.01	0.03	0.01	0.00	0.02	0.06	0.01	0.83	0.82	0.83	0.83
Ajedrez	15	2	(0.12)	(0.01)	(0.04)	(0.06)	(0.04)	(0.00)	(0.01)	(0.02)	(0.02)	(0.12)	(0.12)	(0.12)	(0.12)
Ajedrez	15	5	0.82	0.00	0.02	0.05	0.00	0.00	0.02	0.06	0.01	0.82	0.83	0.83	0.83
Ajedrez	15	5	(0.12)	(0.00)	(0.05)	(0.07)	(0.02)	(0.00)	(0.01)	(0.02)	(0.02)	(0.12)	(0.12)	(0.12)	(0.12)
Ajedrez	30	2	0.83	0.00	0.02	0.04	0.00	0.00	0.03	0.06	0.01	0.83	0.82	0.82	0.83
Ajedrez	30	2	(0.12)	(0.00)	(0.03)	(0.04)	(0.02)	(0.00)	(0.02)	(0.02)	(0.02)	(0.12)	(0.12)	(0.12)	(0.12)
Ajedrez	30	5	0.80	0.00	0.02	0.04	0.01	0.00	0.02	0.06	0.01	0.81	0.81	0.81	0.80
Ajedrez	30	5	(0.13)	(0.00)	(0.03)	(0.05)	(0.02)	(0.00)	(0.01)	(0.02)	(0.02)	(0.13)	(0.14)	(0.13)	(0.14)

**Tabla A-18.:** Promedio y desviación estándar (en paréntesis) de  $ABC$ , tasa de correcta detección ( $v_c$ ), tasa de falsa detección ( $f_c$ ) de cada método de identificación de influentes, en 100 replicaciones. Modelo de respuesta binaria,  $n = 50$ , bajo niveles de contaminación  $\gamma_c$  y factor de contaminación de magnitud  $\gamma_b$ .

$\beta_0$	$\gamma_c(\%)$	$\gamma_b$	ABC	$v_c$				$f_c$				ABC (sin influentes)			
				DO	DO <sub>b</sub> (0.995)	DO <sub>b</sub> (0.95)	MS	DO	DO <sub>b</sub> (0.995)	DO <sub>b</sub> (0.95)	MS	DO	DO <sub>b</sub> (0.995)	DO <sub>b</sub> (0.95)	MS
T	0	1	0.79	-	-	-	-	0.00	0.02	0.05	0.02	0.79	0.79	0.79	0.79
T	0	1	(0.10)	-	-	-	-	(0.00)	(0.01)	(0.01)	(0.02)	(0.10)	(0.11)	(0.10)	(0.10)
T	5	2	0.80	0.00	0.01	0.04	0.01	0.00	0.02	0.05	0.01	0.80	0.80	0.80	0.79
T	5	2	(0.09)	(0.00)	(0.05)	(0.11)	(0.06)	(0.00)	(0.01)	(0.01)	(0.02)	(0.09)	(0.09)	(0.09)	(0.10)
T	5	5	0.79	0.00	0.01	0.04	0.00	0.00	0.02	0.05	0.02	0.79	0.79	0.79	0.79
T	5	5	(0.11)	(0.00)	(0.06)	(0.10)	(0.00)	(0.00)	(0.01)	(0.01)	(0.02)	(0.11)	(0.11)	(0.11)	(0.11)
T	15	2	0.79	0.00	0.01	0.05	0.01	0.00	0.02	0.05	0.02	0.79	0.80	0.80	0.79
T	15	2	(0.11)	(0.00)	(0.03)	(0.07)	(0.03)	(0.00)	(0.01)	(0.02)	(0.03)	(0.11)	(0.11)	(0.11)	(0.11)
T	15	5	0.81	0.00	0.01	0.05	0.01	0.00	0.02	0.05	0.01	0.81	0.82	0.81	0.81
T	15	5	(0.09)	(0.00)	(0.03)	(0.06)	(0.04)	(0.00)	(0.01)	(0.01)	(0.02)	(0.09)	(0.09)	(0.09)	(0.10)
T	30	2	0.79	0.00	0.02	0.05	0.01	0.00	0.02	0.05	0.01	0.79	0.79	0.79	0.79
T	30	2	(0.10)	(0.00)	(0.03)	(0.04)	(0.02)	(0.00)	(0.01)	(0.02)	(0.02)	(0.10)	(0.09)	(0.10)	(0.10)
T	30	5	0.81	0.00	0.01	0.04	0.01	0.00	0.02	0.06	0.02	0.81	0.81	0.81	0.81
T	30	5	(0.11)	(0.00)	(0.02)	(0.04)	(0.03)	(0.00)	(0.01)	(0.02)	(0.02)	(0.11)	(0.11)	(0.11)	(0.11)
Triangulo	0	1	0.73	-	-	-	-	0.00	0.02	0.05	0.01	0.73	0.73	0.72	0.74
Triangulo	0	1	(0.11)	-	-	-	-	(0.00)	(0.00)	(0.01)	(0.01)	(0.11)	(0.11)	(0.12)	(0.11)
Triangulo	5	2	0.75	0.00	0.01	0.06	0.01	0.00	0.02	0.05	0.01	0.75	0.74	0.75	0.75
Triangulo	5	2	(0.12)	(0.00)	(0.07)	(0.13)	(0.05)	(0.00)	(0.01)	(0.01)	(0.01)	(0.12)	(0.12)	(0.12)	(0.12)
Triangulo	5	5	0.75	0.00	0.01	0.04	0.01	0.00	0.02	0.05	0.01	0.75	0.75	0.74	0.75
Triangulo	5	5	(0.11)	(0.00)	(0.06)	(0.10)	(0.05)	(0.00)	(0.01)	(0.01)	(0.02)	(0.11)	(0.11)	(0.11)	(0.11)
Triangulo	15	2	0.74	0.00	0.01	0.03	0.01	0.00	0.02	0.06	0.02	0.74	0.75	0.76	0.75
Triangulo	15	2	(0.11)	(0.00)	(0.02)	(0.05)	(0.02)	(0.00)	(0.01)	(0.01)	(0.02)	(0.11)	(0.11)	(0.11)	(0.11)
Triangulo	15	5	0.76	0.00	0.00	0.03	0.01	0.00	0.02	0.06	0.02	0.76	0.76	0.77	0.76
Triangulo	15	5	(0.12)	(0.00)	(0.02)	(0.05)	(0.03)	(0.00)	(0.01)	(0.01)	(0.03)	(0.12)	(0.12)	(0.12)	(0.12)
Triangulo	30	2	0.75	0.00	0.01	0.04	0.01	0.00	0.02	0.06	0.02	0.75	0.75	0.75	0.75
Triangulo	30	2	(0.11)	(0.00)	(0.02)	(0.04)	(0.02)	(0.00)	(0.01)	(0.02)	(0.02)	(0.11)	(0.11)	(0.10)	(0.11)
Triangulo	30	5	0.75	0.00	0.01	0.04	0.01	0.00	0.02	0.06	0.01	0.75	0.75	0.76	0.75
Triangulo	30	5	(0.10)	(0.00)	(0.02)	(0.04)	(0.02)	(0.00)	(0.01)	(0.02)	(0.01)	(0.10)	(0.10)	(0.10)	(0.10)
Cuadrado	0	1	0.81	-	-	-	-	0.00	0.02	0.05	0.02	0.81	0.81	0.81	0.81
Cuadrado	0	1	(0.10)	-	-	-	-	(0.00)	(0.01)	(0.01)	(0.02)	(0.10)	(0.10)	(0.10)	(0.10)
Cuadrado	5	2	0.77	0.00	0.02	0.04	0.01	0.00	0.02	0.05	0.02	0.77	0.77	0.76	0.77
Cuadrado	5	2	(0.11)	(0.00)	(0.08)	(0.10)	(0.06)	(0.00)	(0.01)	(0.01)	(0.02)	(0.11)	(0.11)	(0.11)	(0.12)
Cuadrado	5	5	0.81	0.00	0.01	0.01	0.01	0.00	0.02	0.06	0.02	0.81	0.81	0.81	0.81
Cuadrado	5	5	(0.10)	(0.00)	(0.05)	(0.07)	(0.06)	(0.00)	(0.01)	(0.01)	(0.03)	(0.10)	(0.10)	(0.09)	(0.09)
Cuadrado	15	2	0.81	0.00	0.02	0.04	0.01	0.00	0.02	0.05	0.02	0.81	0.80	0.81	0.80
Cuadrado	15	2	(0.09)	(0.00)	(0.04)	(0.06)	(0.04)	(0.00)	(0.01)	(0.01)	(0.02)	(0.09)	(0.09)	(0.09)	(0.09)
Cuadrado	15	5	0.81	0.00	0.01	0.04	0.01	0.00	0.02	0.06	0.02	0.81	0.82	0.82	0.81
Cuadrado	15	5	(0.10)	(0.00)	(0.03)	(0.06)	(0.04)	(0.00)	(0.01)	(0.01)	(0.02)	(0.10)	(0.10)	(0.10)	(0.11)
Cuadrado	30	2	0.81	0.00	0.01	0.05	0.01	0.00	0.02	0.06	0.02	0.81	0.82	0.82	0.82
Cuadrado	30	2	(0.11)	(0.00)	(0.02)	(0.04)	(0.02)	(0.00)	(0.01)	(0.02)	(0.02)	(0.11)	(0.10)	(0.09)	(0.11)
Cuadrado	30	5	0.81	0.00	0.01	0.04	0.01	0.00	0.02	0.06	0.02	0.81	0.81	0.82	0.81
Cuadrado	30	5	(0.10)	(0.00)	(0.02)	(0.04)	(0.02)	(0.00)	(0.01)	(0.02)	(0.02)	(0.10)	(0.10)	(0.10)	(0.10)
Ajedrez	0	1	0.84	-	-	-	-	0.00	0.02	0.05	0.02	0.84	0.83	0.83	0.84
Ajedrez	0	1	(0.11)	-	-	-	-	(0.00)	(0.01)	(0.01)	(0.03)	(0.10)	(0.10)	(0.10)	(0.10)
Ajedrez	5	2	0.83	0.00	0.01	0.03	0.01	0.00	0.02	0.05	0.02	0.83	0.83	0.82	0.83
Ajedrez	5	2	(0.10)	(0.00)	(0.07)	(0.10)	(0.06)	(0.00)	(0.01)	(0.01)	(0.02)	(0.10)	(0.10)	(0.09)	(0.09)
Ajedrez	5	5	0.83	0.00	0.01	0.02	0.01	0.00	0.02	0.06	0.02	0.83	0.84	0.84	0.84
Ajedrez	5	5	(0.09)	(0.00)	(0.05)	(0.09)	(0.06)	(0.00)	(0.01)	(0.01)	(0.02)	(0.09)	(0.09)	(0.10)	(0.09)
Ajedrez	15	2	0.83	0.00	0.01	0.05	0.02	0.00	0.02	0.05	0.02	0.83	0.83	0.83	0.83
Ajedrez	15	2	(0.08)	(0.00)	(0.03)	(0.06)	(0.04)	(0.00)	(0.01)	(0.01)	(0.03)	(0.08)	(0.08)	(0.09)	(0.09)
Ajedrez	15	5	0.83	0.00	0.01	0.05	0.01	0.00	0.02	0.05	0.01	0.83	0.83	0.83	0.83
Ajedrez	15	5	(0.10)	(0.00)	(0.03)	(0.06)	(0.03)	(0.00)	(0.01)	(0.01)	(0.03)	(0.10)	(0.10)	(0.09)	(0.10)
Ajedrez	30	2	0.86	0.00	0.01	0.04	0.02	0.00	0.02	0.06	0.02	0.86	0.86	0.86	0.86
Ajedrez	30	2	(0.08)	(0.00)	(0.02)	(0.04)	(0.03)	(0.00)	(0.01)	(0.02)	(0.03)	(0.08)	(0.08)	(0.09)	(0.08)
Ajedrez	30	5	0.86	0.00	0.01	0.05	0.01	0.00	0.02	0.05	0.02	0.86	0.85	0.84	0.86
Ajedrez	30	5	(0.08)	(0.00)	(0.02)	(0.04)	(0.03)	(0.00)	(0.01)	(0.02)	(0.03)	(0.08)	(0.09)	(0.10)	(0.08)

**Tabla A-19.:** Promedio y desviación estándar (en paréntesis) de  $ABC$ , tasa de correcta detección ( $v_c$ ), tasa de falsa detección ( $f_c$ ) de cada método de identificación de influentes, en 100 replicaciones. Modelo de respuesta binaria,  $n = 70$ , bajo niveles de contaminación  $\gamma_c$  y factor de contaminación de magnitud  $\gamma_b$ .

$\beta_0$	$\beta_1$	$\gamma_c(\%)$	$RMSPE$	$v_c$				$f_c$				$RMSPE$ (sin influyentes)			
				$DO$	$DO_b(0.995)$	$DO_b(0.95)$	$MS$	$DO$	$DO_b(0.995)$	$DO_b(0.95)$	$MS$	$DO$	$DO_b(0.995)$	$DO_b(0.95)$	$MS$
T	Tr	5	2.35	0.00	0.00	0.06	0.08	0.00	0.02	0.06	0.11	2.36	2.46	2.59	2.70
T	Tr	5	(0.47)	(0.00)	(0.05)	(0.16)	(0.18)	(0.01)	(0.01)	(0.01)	(0.07)	(0.47)	(0.50)	(0.55)	(0.60)
T	Tr	15	2.41	0.00	0.01	0.04	0.09	0.00	0.02	0.06	0.11	2.42	2.53	2.62	2.76
T	Tr	15	(0.45)	(0.01)	(0.04)	(0.07)	(0.12)	(0.01)	(0.01)	(0.02)	(0.07)	(0.45)	(0.49)	(0.49)	(0.52)
T	Tr	30	2.56	0.00	0.01	0.03	0.06	0.00	0.03	0.07	0.14	2.56	2.69	2.77	2.88
T	Tr	30	(0.45)	(0.00)	(0.02)	(0.04)	(0.07)	(0.00)	(0.01)	(0.02)	(0.07)	(0.45)	(0.46)	(0.47)	(0.47)
T	C	5	2.36	0.00	0.04	0.09	0.15	0.00	0.02	0.05	0.11	2.36	2.47	2.54	2.69
T	C	5	(0.51)	(0.00)	(0.13)	(0.20)	(0.25)	(0.01)	(0.01)	(0.01)	(0.06)	(0.50)	(0.53)	(0.54)	(0.58)
T	C	15	2.52	0.00	0.03	0.06	0.14	0.00	0.02	0.05	0.11	2.53	2.60	2.66	2.80
T	C	15	(0.49)	(0.00)	(0.06)	(0.09)	(0.14)	(0.00)	(0.01)	(0.02)	(0.07)	(0.49)	(0.50)	(0.53)	(0.59)
T	C	30	2.57	0.00	0.02	0.05	0.10	0.00	0.02	0.05	0.11	2.57	2.63	2.66	2.75
T	C	30	(0.41)	(0.00)	(0.04)	(0.05)	(0.09)	(0.00)	(0.02)	(0.02)	(0.08)	(0.42)	(0.39)	(0.41)	(0.41)
T	A	5	2.36	0.00	0.16	0.22	0.28	0.00	0.02	0.05	0.11	2.36	2.42	2.53	2.67
T	A	5	(0.45)	(0.05)	(0.23)	(0.27)	(0.30)	(0.00)	(0.01)	(0.02)	(0.06)	(0.45)	(0.48)	(0.50)	(0.53)
T	A	15	2.54	0.01	0.08	0.14	0.25	0.00	0.01	0.04	0.09	2.54	2.58	2.65	2.76
T	A	15	(0.52)	(0.04)	(0.08)	(0.11)	(0.15)	(0.00)	(0.01)	(0.02)	(0.06)	(0.53)	(0.51)	(0.55)	(0.57)
T	A	30	2.60	0.00	0.05	0.10	0.21	0.00	0.01	0.03	0.09	2.60	2.63	2.65	2.72
T	A	30	(0.43)	(0.01)	(0.04)	(0.05)	(0.11)	(0.00)	(0.01)	(0.02)	(0.06)	(0.43)	(0.43)	(0.45)	(0.48)
Triangulo	T	5	1.90	0.01	0.15	0.23	0.29	0.00	0.02	0.05	0.10	1.90	1.94	1.96	2.03
Triangulo	T	5	(0.38)	(0.09)	(0.24)	(0.30)	(0.35)	(0.00)	(0.01)	(0.02)	(0.06)	(0.38)	(0.36)	(0.37)	(0.38)
Triangulo	T	15	1.97	0.01	0.10	0.19	0.31	0.00	0.01	0.03	0.09	1.97	2.00	2.02	2.08
Triangulo	T	15	(0.34)	(0.03)	(0.06)	(0.11)	(0.18)	(0.00)	(0.01)	(0.02)	(0.06)	(0.34)	(0.36)	(0.36)	(0.36)
Triangulo	T	30	2.01	0.00	0.06	0.12	0.22	0.00	0.01	0.02	0.08	2.01	2.02	2.04	2.10
Triangulo	T	30	(0.40)	(0.01)	(0.03)	(0.04)	(0.11)	(0.00)	(0.01)	(0.02)	(0.06)	(0.40)	(0.38)	(0.38)	(0.39)
Triangulo	C	5	1.85	0.00	0.07	0.12	0.19	0.00	0.02	0.05	0.11	1.85	1.90	1.95	2.03
Triangulo	C	5	(0.32)	(0.00)	(0.17)	(0.21)	(0.27)	(0.00)	(0.01)	(0.02)	(0.06)	(0.32)	(0.32)	(0.34)	(0.37)
Triangulo	C	15	1.91	0.00	0.08	0.13	0.20	0.00	0.01	0.04	0.10	1.92	1.95	1.98	2.05
Triangulo	C	15	(0.34)	(0.02)	(0.08)	(0.10)	(0.13)	(0.00)	(0.01)	(0.02)	(0.06)	(0.35)	(0.37)	(0.37)	(0.34)
Triangulo	C	30	1.98	0.00	0.05	0.10	0.19	0.00	0.01	0.03	0.09	1.98	2.03	2.06	2.11
Triangulo	C	30	(0.43)	(0.00)	(0.04)	(0.05)	(0.12)	(0.00)	(0.01)	(0.02)	(0.06)	(0.43)	(0.46)	(0.48)	(0.49)
Triangulo	A	5	1.88	0.02	0.23	0.32	0.38	0.00	0.01	0.04	0.10	1.88	1.92	1.93	2.01
Triangulo	A	5	(0.33)	(0.10)	(0.27)	(0.29)	(0.35)	(0.00)	(0.01)	(0.02)	(0.06)	(0.33)	(0.35)	(0.36)	(0.34)
Triangulo	A	15	2.02	0.01	0.13	0.25	0.41	0.00	0.01	0.02	0.09	2.01	1.95	1.97	2.03
Triangulo	A	15	(0.39)	(0.03)	(0.07)	(0.11)	(0.17)	(0.00)	(0.01)	(0.02)	(0.05)	(0.39)	(0.37)	(0.38)	(0.37)
Triangulo	A	30	2.24	0.00	0.06	0.14	0.30	0.00	0.00	0.02	0.08	2.23	2.18	2.12	2.14
Triangulo	A	30	(0.44)	(0.02)	(0.03)	(0.05)	(0.12)	(0.00)	(0.01)	(0.02)	(0.06)	(0.42)	(0.40)	(0.43)	(0.41)
Cuadrado	T	5	2.01	0.00	0.09	0.16	0.23	0.00	0.02	0.05	0.10	2.01	2.13	2.21	2.35
Cuadrado	T	5	(0.40)	(0.05)	(0.19)	(0.27)	(0.35)	(0.00)	(0.01)	(0.01)	(0.06)	(0.40)	(0.41)	(0.44)	(0.48)
Cuadrado	T	15	2.12	0.00	0.06	0.13	0.25	0.00	0.02	0.04	0.11	2.13	2.21	2.28	2.44
Cuadrado	T	15	(0.41)	(0.01)	(0.07)	(0.10)	(0.17)	(0.01)	(0.01)	(0.02)	(0.06)	(0.41)	(0.41)	(0.44)	(0.47)
Cuadrado	T	30	2.27	0.00	0.05	0.10	0.17	0.00	0.01	0.03	0.09	2.27	2.31	2.36	2.47
Cuadrado	T	30	(0.44)	(0.01)	(0.03)	(0.06)	(0.11)	(0.00)	(0.01)	(0.02)	(0.06)	(0.43)	(0.46)	(0.47)	(0.51)
Cuadrado	Tr	5	1.99	0.00	0.01	0.06	0.08	0.00	0.02	0.05	0.11	2.00	2.12	2.20	2.32
Cuadrado	Tr	5	(0.40)	(0.00)	(0.09)	(0.17)	(0.21)	(0.01)	(0.01)	(0.02)	(0.06)	(0.42)	(0.45)	(0.44)	(0.47)
Cuadrado	Tr	15	2.07	0.00	0.02	0.03	0.07	0.00	0.02	0.06	0.12	2.08	2.18	2.26	2.40
Cuadrado	Tr	15	(0.45)	(0.00)	(0.05)	(0.07)	(0.10)	(0.00)	(0.01)	(0.02)	(0.07)	(0.45)	(0.46)	(0.50)	(0.49)
Cuadrado	Tr	30	2.13	0.00	0.02	0.04	0.09	0.00	0.02	0.06	0.12	2.13	2.21	2.28	2.37
Cuadrado	Tr	30	(0.46)	(0.01)	(0.03)	(0.05)	(0.08)	(0.00)	(0.01)	(0.03)	(0.07)	(0.46)	(0.51)	(0.52)	(0.56)
Cuadrado	A	5	2.07	0.03	0.17	0.27	0.37	0.00	0.02	0.05	0.10	2.08	2.15	2.23	2.36
Cuadrado	A	5	(0.45)	(0.11)	(0.25)	(0.33)	(0.35)	(0.00)	(0.01)	(0.02)	(0.06)	(0.45)	(0.46)	(0.47)	(0.48)
Cuadrado	A	15	2.21	0.02	0.11	0.21	0.32	0.00	0.01	0.03	0.09	2.21	2.21	2.22	2.37
Cuadrado	A	15	(0.43)	(0.05)	(0.08)	(0.12)	(0.18)	(0.00)	(0.01)	(0.02)	(0.06)	(0.44)	(0.44)	(0.46)	(0.48)
Cuadrado	A	30	2.44	0.00	0.06	0.13	0.23	0.00	0.01	0.02	0.08	2.44	2.41	2.40	2.46
Cuadrado	A	30	(0.49)	(0.01)	(0.03)	(0.06)	(0.11)	(0.00)	(0.01)	(0.02)	(0.06)	(0.49)	(0.49)	(0.49)	(0.52)
Ajedrez	T	5	2.65	0.00	0.04	0.08	0.14	0.00	0.02	0.06	0.11	2.66	2.80	2.89	3.03
Ajedrez	T	5	(0.61)	(0.00)	(0.14)	(0.18)	(0.26)	(0.00)	(0.01)	(0.02)	(0.06)	(0.61)	(0.61)	(0.65)	(0.66)
Ajedrez	T	15	2.87	0.00	0.03	0.07	0.12	0.00	0.02	0.05	0.09	2.87	3.00	3.07	3.21
Ajedrez	T	15	(0.51)	(0.01)	(0.06)	(0.08)	(0.14)	(0.01)	(0.01)	(0.02)	(0.06)	(0.51)	(0.53)	(0.53)	(0.56)
Ajedrez	T	30	3.13	0.00	0.03	0.06	0.10	0.00	0.02	0.05	0.11	3.13	3.15	3.23	3.31
Ajedrez	T	30	(0.60)	(0.00)	(0.03)	(0.05)	(0.08)	(0.00)	(0.02)	(0.02)	(0.07)	(0.60)	(0.63)	(0.61)	(0.64)
Ajedrez	Tr	5	2.88	0.00	0.01	0.06	0.12	0.00	0.03	0.06	0.10	2.89	3.05	3.16	3.36
Ajedrez	Tr	5	(0.57)	(0.00)	(0.07)	(0.17)	(0.23)	(0.00)	(0.01)	(0.02)	(0.06)	(0.57)	(0.58)	(0.58)	(0.66)
Ajedrez	Tr	15	2.85	0.00	0.02	0.06	0.10	0.00	0.02	0.05	0.11	2.85	2.98	3.08	3.22
Ajedrez	Tr	15	(0.57)	(0.00)	(0.05)	(0.08)	(0.12)	(0.00)	(0.01)	(0.02)	(0.08)	(0.57)	(0.59)	(0.62)	(0.62)
Ajedrez	Tr	30	3.19	0.00	0.01	0.03	0.06	0.00	0.03	0.06	0.12	3.20	3.32	3.37	3.47
Ajedrez	Tr	30	(0.71)	(0.00)	(0.02)	(0.04)	(0.07)	(0.00)	(0.01)	(0.02)	(0.07)	(0.71)	(0.70)	(0.70)	(0.75)
Ajedrez	C	5	2.74	0.00	0.02	0.07	0.17	0.00	0.02	0.05	0.10	2.75	2.89	2.99	3.10
Ajedrez	C	5	(0.56)	(0.00)	(0.10)	(0.18)	(0.27)	(0.01)	(0.01)	(0.01)	(0.07)	(0.55)	(0.58)	(0.62)	(0.67)
Ajedrez	C	15	2.84	0.00	0.03	0.08	0.14	0.00	0.02	0.05	0.10	2.84	2.90	3.02	3.16
Ajedrez	C	15	(0.58)	(0.03)	(0.06)	(0.08)	(0.14)	(0.01)	(0.01)	(0.02)	(0.07)	(0.58)	(0.62)	(0.64)	(0.65)
Ajedrez	C	30	3.14	0.00	0.01	0.04	0.10	0.00	0.03	0.06	0.13	3.15	3.25	3.32	3.41
Ajedrez	C	30	(0.62)	(0.00)	(0.02)	(0.04)	(0.10)	(0.01)	(0.01)	(0.02)	(0.08)	(0.63)	(0.67)	(0.67)	(0.69)

**Tabla A-20.:** Promedio y desviación estándar (en paréntesis) de  $RMSPE$ , tasa de correcta detección ( $v_c$ ), tasa de falsa detección ( $f_c$ ) de cada método de identificación de influyentes, en 100 replicaciones. Modelo de respuesta continua,  $n = 50$ , bajo de contaminación de cambio forma - Tipo 1 en parámetro imagen.

$\beta_0$	$\beta_1$	$\gamma_c(\%)$	$RMSPE$	$v_c$				$f_c$				$RMSPE$ (sin influentes)			
				$DO$	$DO_b(0.995)$	$DO_b(0.95)$	$MS$	$DO$	$DO_b(0.995)$	$DO_b(0.95)$	$MS$	$DO$	$DO_b(0.995)$	$DO_b(0.95)$	$MS$
T	Tr	5	2.14	0.00	0.03	0.10	0.16	0.00	0.02	0.05	0.11	2.15	2.21	2.29	2.44
T	Tr	5	(0.39)	(0.03)	(0.10)	(0.17)	(0.19)	(0.01)	(0.01)	(0.01)	(0.05)	(0.38)	(0.41)	(0.41)	(0.46)
T	Tr	15	2.33	0.00	0.02	0.05	0.12	0.00	0.02	0.05	0.11	2.34	2.43	2.51	2.65
T	Tr	15	(0.41)	(0.00)	(0.04)	(0.06)	(0.10)	(0.00)	(0.01)	(0.01)	(0.06)	(0.41)	(0.44)	(0.45)	(0.44)
T	Tr	30	2.42	0.00	0.01	0.04	0.09	0.00	0.02	0.06	0.13	2.42	2.47	2.53	2.67
T	Tr	30	(0.40)	(0.00)	(0.02)	(0.04)	(0.07)	(0.01)	(0.01)	(0.02)	(0.07)	(0.40)	(0.41)	(0.43)	(0.43)
T	C	5	2.18	0.00	0.02	0.09	0.17	0.00	0.02	0.05	0.11	2.20	2.27	2.39	2.53
T	C	5	(0.33)	(0.00)	(0.09)	(0.17)	(0.22)	(0.01)	(0.00)	(0.01)	(0.06)	(0.34)	(0.36)	(0.39)	(0.43)
T	C	15	2.33	0.00	0.03	0.07	0.15	0.00	0.01	0.05	0.13	2.33	2.38	2.45	2.59
T	C	15	(0.38)	(0.00)	(0.04)	(0.06)	(0.12)	(0.00)	(0.01)	(0.02)	(0.06)	(0.38)	(0.39)	(0.41)	(0.44)
T	C	30	2.33	0.00	0.02	0.06	0.12	0.00	0.02	0.05	0.12	2.34	2.39	2.44	2.54
T	C	30	(0.43)	(0.00)	(0.02)	(0.04)	(0.07)	(0.00)	(0.01)	(0.02)	(0.05)	(0.43)	(0.44)	(0.44)	(0.46)
T	A	5	2.19	0.02	0.09	0.21	0.32	0.00	0.01	0.04	0.11	2.20	2.25	2.35	2.50
T	A	5	(0.32)	(0.09)	(0.15)	(0.22)	(0.28)	(0.00)	(0.01)	(0.01)	(0.06)	(0.34)	(0.38)	(0.41)	(0.44)
T	A	15	2.42	0.00	0.07	0.17	0.30	0.00	0.01	0.03	0.10	2.42	2.40	2.47	2.59
T	A	15	(0.39)	(0.01)	(0.05)	(0.08)	(0.14)	(0.00)	(0.01)	(0.01)	(0.05)	(0.39)	(0.39)	(0.38)	(0.38)
T	A	30	2.68	0.00	0.04	0.11	0.23	0.00	0.01	0.03	0.09	2.68	2.66	2.67	2.72
T	A	30	(0.41)	(0.01)	(0.02)	(0.04)	(0.08)	(0.00)	(0.01)	(0.02)	(0.05)	(0.40)	(0.40)	(0.39)	(0.44)
Triangulo	T	5	1.73	0.01	0.13	0.26	0.42	0.00	0.01	0.04	0.12	1.73	1.76	1.82	1.90
Triangulo	T	5	(0.28)	(0.05)	(0.16)	(0.23)	(0.28)	(0.00)	(0.01)	(0.01)	(0.06)	(0.28)	(0.29)	(0.32)	(0.31)
Triangulo	T	15	1.80	0.00	0.07	0.17	0.30	0.00	0.01	0.03	0.11	1.80	1.82	1.87	1.94
Triangulo	T	15	(0.26)	(0.02)	(0.05)	(0.09)	(0.16)	(0.00)	(0.01)	(0.02)	(0.06)	(0.26)	(0.26)	(0.27)	(0.30)
Triangulo	T	30	1.95	0.00	0.04	0.12	0.23	0.00	0.01	0.03	0.09	1.95	1.95	1.97	2.00
Triangulo	T	30	(0.38)	(0.01)	(0.02)	(0.05)	(0.10)	(0.00)	(0.01)	(0.02)	(0.05)	(0.38)	(0.36)	(0.37)	(0.37)
Triangulo	C	5	1.73	0.02	0.10	0.19	0.28	0.00	0.01	0.05	0.10	1.73	1.78	1.81	1.89
Triangulo	C	5	(0.28)	(0.07)	(0.15)	(0.21)	(0.26)	(0.00)	(0.01)	(0.01)	(0.06)	(0.28)	(0.28)	(0.28)	(0.28)
Triangulo	C	15	1.86	0.00	0.06	0.13	0.22	0.00	0.01	0.04	0.11	1.86	1.88	1.91	2.00
Triangulo	C	15	(0.31)	(0.02)	(0.06)	(0.09)	(0.13)	(0.00)	(0.01)	(0.02)	(0.06)	(0.31)	(0.31)	(0.31)	(0.34)
Triangulo	C	30	1.82	0.00	0.04	0.10	0.20	0.00	0.01	0.03	0.10	1.82	1.84	1.84	1.90
Triangulo	C	30	(0.32)	(0.00)	(0.03)	(0.05)	(0.09)	(0.00)	(0.01)	(0.02)	(0.06)	(0.32)	(0.31)	(0.30)	(0.35)
Triangulo	A	5	1.80	0.05	0.20	0.36	0.42	0.00	0.01	0.04	0.12	1.78	1.78	1.86	1.95
Triangulo	A	5	(0.32)	(0.13)	(0.16)	(0.25)	(0.28)	(0.00)	(0.01)	(0.02)	(0.05)	(0.31)	(0.30)	(0.30)	(0.32)
Triangulo	A	15	2.00	0.02	0.10	0.24	0.40	0.00	0.00	0.02	0.10	1.97	1.92	1.92	1.96
Triangulo	A	15	(0.32)	(0.05)	(0.04)	(0.08)	(0.13)	(0.00)	(0.01)	(0.01)	(0.05)	(0.32)	(0.30)	(0.31)	(0.33)
Triangulo	A	30	2.25	0.00	0.05	0.14	0.27	0.00	0.00	0.01	0.07	2.24	2.14	2.09	2.09
Triangulo	A	30	(0.38)	(0.01)	(0.02)	(0.04)	(0.10)	(0.00)	(0.00)	(0.01)	(0.05)	(0.37)	(0.37)	(0.32)	(0.33)
Cuadrado	T	5	1.88	0.00	0.10	0.19	0.27	0.00	0.01	0.05	0.11	1.89	1.94	2.06	2.21
Cuadrado	T	5	(0.35)	(0.00)	(0.18)	(0.24)	(0.27)	(0.00)	(0.01)	(0.02)	(0.05)	(0.35)	(0.37)	(0.37)	(0.42)
Cuadrado	T	15	1.99	0.01	0.06	0.14	0.24	0.00	0.01	0.04	0.11	1.99	2.01	2.09	2.22
Cuadrado	T	15	(0.36)	(0.03)	(0.06)	(0.09)	(0.14)	(0.00)	(0.01)	(0.02)	(0.06)	(0.35)	(0.35)	(0.35)	(0.38)
Cuadrado	T	30	2.09	0.00	0.04	0.10	0.19	0.00	0.01	0.04	0.10	2.09	2.09	2.16	2.28
Cuadrado	T	30	(0.34)	(0.01)	(0.02)	(0.05)	(0.08)	(0.00)	(0.01)	(0.02)	(0.05)	(0.35)	(0.34)	(0.34)	(0.38)
Cuadrado	Tr	5	1.79	0.00	0.04	0.09	0.19	0.00	0.01	0.05	0.13	1.80	1.86	1.96	2.15
Cuadrado	Tr	5	(0.32)	(0.03)	(0.10)	(0.17)	(0.23)	(0.01)	(0.01)	(0.01)	(0.06)	(0.32)	(0.33)	(0.34)	(0.35)
Cuadrado	Tr	15	1.91	0.00	0.01	0.05	0.10	0.00	0.02	0.05	0.12	1.92	1.99	2.05	2.18
Cuadrado	Tr	15	(0.34)	(0.00)	(0.04)	(0.07)	(0.10)	(0.01)	(0.01)	(0.02)	(0.06)	(0.34)	(0.34)	(0.38)	(0.40)
Cuadrado	Tr	30	2.05	0.00	0.01	0.04	0.10	0.00	0.02	0.06	0.13	2.05	2.13	2.21	2.32
Cuadrado	Tr	30	(0.35)	(0.00)	(0.02)	(0.04)	(0.08)	(0.00)	(0.01)	(0.02)	(0.06)	(0.35)	(0.36)	(0.38)	(0.39)
Cuadrado	A	5	1.91	0.03	0.20	0.35	0.44	0.00	0.01	0.04	0.10	1.91	1.90	1.99	2.13
Cuadrado	A	5	(0.32)	(0.11)	(0.18)	(0.23)	(0.25)	(0.00)	(0.01)	(0.01)	(0.05)	(0.32)	(0.32)	(0.33)	(0.37)
Cuadrado	A	15	2.12	0.02	0.09	0.21	0.33	0.00	0.00	0.03	0.09	2.11	2.07	2.09	2.22
Cuadrado	A	15	(0.37)	(0.04)	(0.04)	(0.08)	(0.13)	(0.01)	(0.01)	(0.02)	(0.05)	(0.38)	(0.38)	(0.38)	(0.41)
Cuadrado	A	30	2.44	0.00	0.05	0.13	0.27	0.00	0.00	0.02	0.08	2.43	2.38	2.37	2.42
Cuadrado	A	30	(0.42)	(0.01)	(0.02)	(0.04)	(0.09)	(0.00)	(0.01)	(0.01)	(0.05)	(0.42)	(0.40)	(0.43)	(0.43)
Ajedrez	T	5	2.48	0.00	0.04	0.11	0.19	0.00	0.02	0.05	0.11	2.49	2.57	2.67	2.89
Ajedrez	T	5	(0.43)	(0.00)	(0.11)	(0.17)	(0.23)	(0.00)	(0.01)	(0.01)	(0.05)	(0.43)	(0.46)	(0.51)	(0.53)
Ajedrez	T	15	2.72	0.00	0.04	0.10	0.20	0.00	0.01	0.04	0.11	2.73	2.80	2.89	3.05
Ajedrez	T	15	(0.43)	(0.02)	(0.05)	(0.09)	(0.13)	(0.00)	(0.01)	(0.02)	(0.06)	(0.44)	(0.49)	(0.51)	(0.57)
Ajedrez	T	30	2.90	0.00	0.02	0.07	0.12	0.00	0.02	0.05	0.12	2.90	2.97	3.05	3.21
Ajedrez	T	30	(0.52)	(0.00)	(0.03)	(0.04)	(0.07)	(0.00)	(0.01)	(0.02)	(0.06)	(0.52)	(0.51)	(0.48)	(0.51)
Ajedrez	Tr	5	2.52	0.00	0.04	0.12	0.22	0.00	0.02	0.05	0.12	2.52	2.61	2.75	2.93
Ajedrez	Tr	5	(0.45)	(0.00)	(0.11)	(0.17)	(0.26)	(0.01)	(0.01)	(0.01)	(0.06)	(0.45)	(0.45)	(0.47)	(0.48)
Ajedrez	Tr	15	2.79	0.00	0.03	0.08	0.16	0.00	0.01	0.05	0.12	2.80	2.84	2.93	3.08
Ajedrez	Tr	15	(0.51)	(0.01)	(0.05)	(0.07)	(0.13)	(0.00)	(0.01)	(0.01)	(0.06)	(0.51)	(0.53)	(0.53)	(0.56)
Ajedrez	Tr	30	3.01	0.00	0.01	0.04	0.10	0.00	0.02	0.06	0.13	3.01	3.11	3.22	3.33
Ajedrez	Tr	30	(0.57)	(0.00)	(0.02)	(0.03)	(0.07)	(0.00)	(0.01)	(0.02)	(0.06)	(0.58)	(0.56)	(0.59)	(0.61)
Ajedrez	C	5	2.44	0.00	0.04	0.10	0.19	0.00	0.01	0.05	0.11	2.44	2.53	2.66	2.86
Ajedrez	C	5	(0.47)	(0.00)	(0.11)	(0.16)	(0.21)	(0.00)	(0.01)	(0.01)	(0.06)	(0.47)	(0.50)	(0.52)	(0.54)
Ajedrez	C	15	2.71	0.00	0.02	0.09	0.16	0.00	0.01	0.05	0.10	2.71	2.78	2.87	3.02
Ajedrez	C	15	(0.47)	(0.00)	(0.04)	(0.08)	(0.12)	(0.00)	(0.01)	(0.01)	(0.06)	(0.47)	(0.46)	(0.51)	(0.54)
Ajedrez	C	30	2.87	0.00	0.01	0.05	0.11	0.00	0.02	0.06	0.12	2.88	2.98	3.05	3.17
Ajedrez	C	30	(0.52)	(0.00)	(0.02)	(0.04)	(0.08)	(0.00)	(0.01)	(0.02)	(0.06)	(0.53)	(0.56)	(0.59)	(0.62)

**Tabla A-21.:** Promedio y desviación estándar (en paréntesis) de  $RMSPE$ , tasa de correcta detección ( $v_c$ ), tasa de falsa detección ( $f_c$ ) de cada método de identificación de influentes, en 100 replicaciones. Modelo de respuesta continua,  $n = 70$ , bajo de contaminación de cambio forma - Tipo 1 en parámetro imagen.

$\beta_0$	$\beta_1$	$\gamma_c(\%)$	$ABC$	$v_c$				$f_c$				$ABC$ (sin influentes)			
				$DO$	$DO_b(0.995)$	$DO_b(0.95)$	$MS$	$DO$	$DO_b(0.995)$	$DO_b(0.95)$	$MS$	$DO$	$DO_b(0.995)$	$DO_b(0.95)$	$MS$
T	Tr	5	0.79	0.00	0.06	0.11	0.02	0.00	0.02	0.05	0.01	0.79	0.78	0.79	0.78
T	Tr	5	(0.13)	(0.00)	(0.16)	(0.22)	(0.12)	(0.00)	(0.01)	(0.02)	(0.02)	(0.13)	(0.13)	(0.13)	(0.13)
T	Tr	15	0.76	0.00	0.03	0.09	0.02	0.00	0.02	0.05	0.01	0.76	0.76	0.77	0.76
T	Tr	15	(0.13)	(0.00)	(0.06)	(0.10)	(0.06)	(0.00)	(0.01)	(0.02)	(0.02)	(0.13)	(0.13)	(0.13)	(0.13)
T	Tr	30	0.75	0.00	0.03	0.07	0.01	0.00	0.02	0.05	0.01	0.75	0.75	0.76	0.74
T	Tr	30	(0.14)	(0.00)	(0.03)	(0.05)	(0.03)	(0.00)	(0.02)	(0.02)	(0.02)	(0.14)	(0.12)	(0.12)	(0.14)
T	C	5	0.78	0.00	0.04	0.09	0.01	0.00	0.02	0.06	0.01	0.78	0.78	0.78	0.78
T	C	5	(0.14)	(0.00)	(0.14)	(0.19)	(0.09)	(0.00)	(0.01)	(0.02)	(0.02)	(0.14)	(0.14)	(0.14)	(0.14)
T	C	15	0.76	0.00	0.04	0.08	0.02	0.00	0.02	0.05	0.01	0.76	0.76	0.76	0.76
T	C	15	(0.13)	(0.00)	(0.06)	(0.10)	(0.05)	(0.00)	(0.01)	(0.02)	(0.02)	(0.13)	(0.13)	(0.15)	(0.13)
T	C	30	0.77	0.00	0.03	0.06	0.01	0.00	0.02	0.05	0.01	0.77	0.78	0.77	0.77
T	C	30	(0.12)	(0.00)	(0.04)	(0.06)	(0.04)	(0.00)	(0.01)	(0.03)	(0.03)	(0.12)	(0.13)	(0.13)	(0.13)
T	A	5	0.78	0.00	0.04	0.08	0.02	0.00	0.02	0.05	0.01	0.78	0.78	0.77	0.78
T	A	5	(0.13)	(0.00)	(0.13)	(0.18)	(0.12)	(0.00)	(0.01)	(0.01)	(0.02)	(0.13)	(0.13)	(0.13)	(0.13)
T	A	15	0.77	0.00	0.03	0.07	0.01	0.00	0.02	0.05	0.01	0.77	0.77	0.78	0.77
T	A	15	(0.13)	(0.01)	(0.06)	(0.09)	(0.03)	(0.00)	(0.01)	(0.02)	(0.02)	(0.13)	(0.12)	(0.12)	(0.13)
T	A	30	0.74	0.00	0.03	0.07	0.01	0.00	0.02	0.05	0.01	0.74	0.73	0.73	0.74
T	A	30	(0.15)	(0.01)	(0.04)	(0.05)	(0.03)	(0.00)	(0.02)	(0.02)	(0.01)	(0.16)	(0.16)	(0.15)	(0.16)
Triangulo	T	5	0.72	0.00	0.01	0.04	0.00	0.00	0.02	0.06	0.01	0.72	0.71	0.72	0.71
Triangulo	T	5	(0.16)	(0.00)	(0.07)	(0.14)	(0.00)	(0.00)	(0.01)	(0.01)	(0.02)	(0.16)	(0.15)	(0.15)	(0.15)
Triangulo	T	15	0.68	0.00	0.02	0.06	0.01	0.00	0.02	0.05	0.01	0.68	0.68	0.68	0.68
Triangulo	T	15	(0.16)	(0.00)	(0.05)	(0.09)	(0.03)	(0.00)	(0.01)	(0.02)	(0.01)	(0.16)	(0.16)	(0.15)	(0.16)
Triangulo	T	30	0.70	0.00	0.03	0.06	0.01	0.00	0.02	0.05	0.01	0.70	0.70	0.70	0.71
Triangulo	T	30	(0.15)	(0.00)	(0.04)	(0.05)	(0.03)	(0.00)	(0.02)	(0.03)	(0.02)	(0.15)	(0.15)	(0.16)	(0.15)
Triangulo	C	5	0.69	0.00	0.01	0.05	0.00	0.00	0.02	0.05	0.01	0.69	0.69	0.70	0.70
Triangulo	C	5	(0.15)	(0.00)	(0.09)	(0.15)	(0.00)	(0.00)	(0.01)	(0.01)	(0.02)	(0.15)	(0.15)	(0.15)	(0.15)
Triangulo	C	15	0.71	0.00	0.02	0.05	0.01	0.00	0.02	0.06	0.01	0.71	0.71	0.71	0.71
Triangulo	C	15	(0.14)	(0.00)	(0.05)	(0.08)	(0.05)	(0.00)	(0.01)	(0.02)	(0.01)	(0.14)	(0.13)	(0.13)	(0.14)
Triangulo	C	30	0.72	0.00	0.03	0.05	0.01	0.00	0.02	0.05	0.01	0.72	0.72	0.72	0.73
Triangulo	C	30	(0.14)	(0.01)	(0.03)	(0.05)	(0.02)	(0.00)	(0.02)	(0.02)	(0.02)	(0.13)	(0.13)	(0.13)	(0.13)
Triangulo	A	5	0.72	0.00	0.07	0.08	0.02	0.00	0.02	0.05	0.01	0.72	0.72	0.72	0.72
Triangulo	A	5	(0.14)	(0.00)	(0.17)	(0.18)	(0.10)	(0.00)	(0.01)	(0.02)	(0.01)	(0.14)	(0.13)	(0.14)	(0.13)
Triangulo	A	15	0.71	0.00	0.03	0.08	0.01	0.00	0.02	0.05	0.01	0.71	0.71	0.72	0.72
Triangulo	A	15	(0.16)	(0.02)	(0.06)	(0.10)	(0.04)	(0.00)	(0.01)	(0.02)	(0.01)	(0.16)	(0.16)	(0.15)	(0.16)
Triangulo	A	30	0.70	0.00	0.03	0.06	0.01	0.00	0.02	0.05	0.01	0.70	0.69	0.69	0.71
Triangulo	A	30	(0.14)	(0.01)	(0.03)	(0.06)	(0.03)	(0.00)	(0.02)	(0.03)	(0.01)	(0.15)	(0.15)	(0.16)	(0.14)
Cuadrado	T	5	0.78	0.00	0.04	0.07	0.01	0.00	0.02	0.05	0.01	0.78	0.79	0.80	0.79
Cuadrado	T	5	(0.14)	(0.05)	(0.14)	(0.18)	(0.09)	(0.00)	(0.01)	(0.02)	(0.01)	(0.14)	(0.14)	(0.13)	(0.13)
Cuadrado	T	15	0.79	0.00	0.03	0.06	0.02	0.00	0.02	0.05	0.01	0.79	0.78	0.78	0.78
Cuadrado	T	15	(0.13)	(0.00)	(0.06)	(0.08)	(0.05)	(0.00)	(0.01)	(0.02)	(0.02)	(0.13)	(0.13)	(0.13)	(0.13)
Cuadrado	T	30	0.77	0.00	0.02	0.06	0.01	0.00	0.02	0.05	0.01	0.77	0.77	0.78	0.77
Cuadrado	T	30	(0.13)	(0.00)	(0.04)	(0.06)	(0.03)	(0.00)	(0.01)	(0.03)	(0.02)	(0.13)	(0.13)	(0.12)	(0.13)
Cuadrado	Tr	5	0.77	0.00	0.04	0.10	0.03	0.00	0.02	0.05	0.01	0.77	0.77	0.77	0.77
Cuadrado	Tr	5	(0.12)	(0.00)	(0.14)	(0.21)	(0.14)	(0.00)	(0.01)	(0.01)	(0.02)	(0.12)	(0.13)	(0.12)	(0.12)
Cuadrado	Tr	15	0.81	0.00	0.04	0.09	0.02	0.00	0.02	0.05	0.01	0.81	0.80	0.80	0.81
Cuadrado	Tr	15	(0.12)	(0.00)	(0.07)	(0.09)	(0.05)	(0.00)	(0.01)	(0.02)	(0.02)	(0.12)	(0.12)	(0.11)	(0.12)
Cuadrado	Tr	30	0.79	0.00	0.03	0.06	0.01	0.00	0.02	0.05	0.01	0.79	0.78	0.78	0.78
Cuadrado	Tr	30	(0.12)	(0.00)	(0.04)	(0.06)	(0.02)	(0.01)	(0.02)	(0.03)	(0.01)	(0.12)	(0.13)	(0.12)	(0.13)
Cuadrado	A	5	0.78	0.00	0.06	0.09	0.03	0.00	0.02	0.05	0.01	0.78	0.79	0.79	0.78
Cuadrado	A	5	(0.12)	(0.00)	(0.18)	(0.22)	(0.11)	(0.00)	(0.01)	(0.02)	(0.02)	(0.12)	(0.11)	(0.11)	(0.12)
Cuadrado	A	15	0.81	0.00	0.03	0.08	0.01	0.00	0.02	0.05	0.01	0.81	0.81	0.81	0.81
Cuadrado	A	15	(0.13)	(0.00)	(0.07)	(0.10)	(0.04)	(0.00)	(0.01)	(0.02)	(0.03)	(0.13)	(0.13)	(0.14)	(0.13)
Cuadrado	A	30	0.73	0.00	0.03	0.07	0.02	0.00	0.02	0.05	0.01	0.73	0.74	0.74	0.74
Cuadrado	A	30	(0.15)	(0.00)	(0.04)	(0.06)	(0.05)	(0.00)	(0.02)	(0.03)	(0.02)	(0.15)	(0.14)	(0.15)	(0.15)
Ajedrez	T	5	0.81	0.00	0.03	0.10	0.04	0.00	0.02	0.05	0.01	0.81	0.83	0.82	0.81
Ajedrez	T	5	(0.12)	(0.00)	(0.12)	(0.20)	(0.14)	(0.00)	(0.01)	(0.01)	(0.03)	(0.12)	(0.11)	(0.11)	(0.12)
Ajedrez	T	15	0.80	0.00	0.04	0.10	0.02	0.00	0.02	0.05	0.01	0.80	0.80	0.81	0.81
Ajedrez	T	15	(0.13)	(0.00)	(0.06)	(0.10)	(0.05)	(0.00)	(0.01)	(0.02)	(0.02)	(0.13)	(0.11)	(0.12)	(0.12)
Ajedrez	T	30	0.77	0.00	0.03	0.07	0.01	0.00	0.02	0.05	0.01	0.77	0.78	0.78	0.78
Ajedrez	T	30	(0.14)	(0.00)	(0.03)	(0.05)	(0.04)	(0.00)	(0.02)	(0.02)	(0.01)	(0.14)	(0.14)	(0.14)	(0.14)
Ajedrez	Tr	5	0.81	0.00	0.09	0.15	0.06	0.00	0.02	0.05	0.01	0.81	0.80	0.80	0.81
Ajedrez	Tr	5	(0.12)	(0.00)	(0.19)	(0.24)	(0.17)	(0.00)	(0.01)	(0.01)	(0.02)	(0.12)	(0.12)	(0.12)	(0.12)
Ajedrez	Tr	15	0.78	0.00	0.05	0.11	0.03	0.00	0.02	0.04	0.01	0.78	0.79	0.80	0.79
Ajedrez	Tr	15	(0.14)	(0.00)	(0.07)	(0.11)	(0.08)	(0.00)	(0.01)	(0.02)	(0.02)	(0.14)	(0.13)	(0.13)	(0.14)
Ajedrez	Tr	30	0.76	0.00	0.04	0.09	0.02	0.00	0.01	0.04	0.00	0.76	0.76	0.76	0.76
Ajedrez	Tr	30	(0.15)	(0.00)	(0.04)	(0.06)	(0.04)	(0.00)	(0.02)	(0.02)	(0.01)	(0.15)	(0.16)	(0.16)	(0.15)
Ajedrez	C	5	0.81	0.00	0.09	0.13	0.04	0.00	0.02	0.05	0.01	0.81	0.81	0.82	0.81
Ajedrez	C	5	(0.12)	(0.00)	(0.19)	(0.22)	(0.14)	(0.00)	(0.01)	(0.02)	(0.02)	(0.12)	(0.12)	(0.12)	(0.12)
Ajedrez	C	15	0.77	0.00	0.05	0.09	0.03	0.00	0.02	0.05	0.01	0.77	0.77	0.78	0.77
Ajedrez	C	15	(0.14)	(0.00)	(0.07)	(0.10)	(0.07)	(0.00)	(0.01)	(0.02)	(0.02)	(0.14)	(0.13)	(0.13)	(0.14)
Ajedrez	C	30	0.78	0.00	0.03	0.08	0.02	0.00	0.02	0.04	0.01	0.78	0.78	0.77	0.78
Ajedrez	C	30	(0.13)	(0.01)	(0.04)	(0.06)	(0.04)	(0.00)	(0.02)	(0.02)	(0.03)	(0.13)	(0.12)	(0.12)	(0.13)

**Tabla A-22.:** Promedio y desviación estándar (en paréntesis) de  $ABC$ , tasa de correcta detección ( $v_c$ ), tasa de falsa detección ( $f_c$ ) de cada método de identificación de influentes, en 100 replicaciones. Modelo de respuesta binaria,  $n = 50$ , bajo de contaminación de cambio forma - Tipo 1 en parámetro imagen.

$\beta_0$	$\beta_1$	$\gamma_c(\%)$	$ABC$	$v_c$				$f_c$				$ABC$ (sin influentes)			
				$DO$	$DO_b(0.995)$	$DO_b(0.95)$	$MS$	$DO$	$DO_b(0.995)$	$DO_b(0.95)$	$MS$	$DO$	$DO_b(0.995)$	$DO_b(0.95)$	$MS$
T	Tr	5	0.81	0.00	0.04	0.11	0.03	0.00	0.02	0.05	0.01	0.81	0.80	0.80	0.80
T	Tr	5	(0.10)	(0.00)	(0.11)	(0.16)	(0.10)	(0.00)	(0.01)	(0.01)	(0.01)	(0.10)	(0.10)	(0.10)	(0.11)
T	Tr	15	0.80	0.00	0.03	0.09	0.03	0.00	0.01	0.05	0.01	0.80	0.79	0.79	0.79
T	Tr	15	(0.11)	(0.00)	(0.05)	(0.07)	(0.05)	(0.00)	(0.01)	(0.01)	(0.02)	(0.10)	(0.10)	(0.11)	(0.10)
T	Tr	30	0.79	0.00	0.03	0.07	0.02	0.00	0.01	0.04	0.01	0.79	0.79	0.79	0.79
T	Tr	30	(0.12)	(0.00)	(0.02)	(0.05)	(0.03)	(0.00)	(0.01)	(0.02)	(0.02)	(0.12)	(0.12)	(0.11)	(0.12)
T	C	5	0.78	0.00	0.04	0.09	0.03	0.00	0.01	0.05	0.01	0.78	0.78	0.79	0.79
T	C	5	(0.10)	(0.00)	(0.11)	(0.17)	(0.10)	(0.00)	(0.01)	(0.01)	(0.02)	(0.10)	(0.10)	(0.11)	(0.10)
T	C	15	0.77	0.00	0.02	0.07	0.02	0.00	0.02	0.05	0.01	0.77	0.77	0.78	0.77
T	C	15	(0.12)	(0.00)	(0.04)	(0.07)	(0.05)	(0.00)	(0.01)	(0.02)	(0.02)	(0.12)	(0.13)	(0.12)	(0.12)
T	C	30	0.78	0.00	0.02	0.06	0.02	0.00	0.02	0.05	0.02	0.78	0.77	0.78	0.78
T	C	30	(0.10)	(0.00)	(0.02)	(0.05)	(0.03)	(0.00)	(0.01)	(0.02)	(0.02)	(0.10)	(0.10)	(0.09)	(0.10)
T	A	5	0.80	0.00	0.06	0.10	0.05	0.00	0.02	0.05	0.01	0.80	0.80	0.80	0.81
T	A	5	(0.11)	(0.00)	(0.14)	(0.17)	(0.13)	(0.00)	(0.01)	(0.01)	(0.02)	(0.11)	(0.11)	(0.12)	(0.11)
T	A	15	0.77	0.00	0.03	0.07	0.03	0.00	0.01	0.05	0.01	0.77	0.78	0.79	0.78
T	A	15	(0.12)	(0.00)	(0.05)	(0.07)	(0.06)	(0.00)	(0.01)	(0.02)	(0.02)	(0.12)	(0.11)	(0.11)	(0.11)
T	A	30	0.76	0.00	0.02	0.05	0.01	0.00	0.02	0.05	0.02	0.76	0.78	0.78	0.77
T	A	30	(0.12)	(0.00)	(0.03)	(0.04)	(0.03)	(0.00)	(0.01)	(0.02)	(0.02)	(0.12)	(0.11)	(0.11)	(0.12)
Triangulo	T	5	0.75	0.00	0.01	0.07	0.02	0.00	0.02	0.05	0.01	0.75	0.74	0.74	0.75
Triangulo	T	5	(0.11)	(0.00)	(0.07)	(0.15)	(0.08)	(0.00)	(0.01)	(0.01)	(0.02)	(0.11)	(0.12)	(0.11)	(0.11)
Triangulo	T	15	0.75	0.00	0.02	0.05	0.01	0.00	0.01	0.05	0.01	0.75	0.74	0.75	0.75
Triangulo	T	15	(0.12)	(0.00)	(0.04)	(0.06)	(0.04)	(0.00)	(0.01)	(0.01)	(0.02)	(0.12)	(0.13)	(0.11)	(0.11)
Triangulo	T	30	0.70	0.00	0.02	0.05	0.01	0.00	0.02	0.05	0.01	0.70	0.71	0.72	0.70
Triangulo	T	30	(0.12)	(0.00)	(0.03)	(0.04)	(0.03)	(0.00)	(0.01)	(0.02)	(0.02)	(0.12)	(0.12)	(0.13)	(0.12)
Triangulo	C	5	0.75	0.00	0.02	0.06	0.01	0.00	0.02	0.05	0.01	0.75	0.75	0.75	0.75
Triangulo	C	5	(0.11)	(0.00)	(0.08)	(0.13)	(0.05)	(0.00)	(0.01)	(0.01)	(0.02)	(0.11)	(0.11)	(0.10)	(0.12)
Triangulo	C	15	0.73	0.00	0.02	0.05	0.01	0.00	0.02	0.05	0.01	0.73	0.73	0.74	0.73
Triangulo	C	15	(0.12)	(0.01)	(0.04)	(0.06)	(0.04)	(0.00)	(0.01)	(0.01)	(0.02)	(0.12)	(0.12)	(0.12)	(0.12)
Triangulo	C	30	0.74	0.00	0.02	0.05	0.01	0.00	0.02	0.05	0.01	0.74	0.73	0.73	0.73
Triangulo	C	30	(0.10)	(0.00)	(0.03)	(0.04)	(0.03)	(0.00)	(0.01)	(0.02)	(0.02)	(0.10)	(0.11)	(0.11)	(0.11)
Triangulo	A	5	0.73	0.00	0.02	0.09	0.01	0.00	0.02	0.05	0.01	0.73	0.74	0.75	0.74
Triangulo	A	5	(0.12)	(0.00)	(0.09)	(0.16)	(0.05)	(0.00)	(0.01)	(0.01)	(0.02)	(0.12)	(0.12)	(0.12)	(0.12)
Triangulo	A	15	0.72	0.00	0.03	0.07	0.01	0.00	0.01	0.05	0.01	0.72	0.73	0.73	0.73
Triangulo	A	15	(0.11)	(0.00)	(0.05)	(0.08)	(0.04)	(0.00)	(0.01)	(0.02)	(0.01)	(0.11)	(0.11)	(0.11)	(0.11)
Triangulo	A	30	0.70	0.00	0.02	0.06	0.01	0.00	0.02	0.05	0.01	0.70	0.70	0.70	0.70
Triangulo	A	30	(0.12)	(0.00)	(0.02)	(0.04)	(0.02)	(0.00)	(0.01)	(0.02)	(0.02)	(0.12)	(0.13)	(0.13)	(0.12)
Cuadrado	T	5	0.79	0.00	0.01	0.06	0.02	0.00	0.02	0.05	0.01	0.79	0.80	0.80	0.79
Cuadrado	T	5	(0.12)	(0.00)	(0.07)	(0.13)	(0.09)	(0.00)	(0.01)	(0.01)	(0.03)	(0.12)	(0.11)	(0.12)	(0.11)
Cuadrado	T	15	0.79	0.00	0.02	0.08	0.02	0.00	0.02	0.05	0.01	0.79	0.79	0.79	0.79
Cuadrado	T	15	(0.11)	(0.00)	(0.04)	(0.07)	(0.05)	(0.00)	(0.01)	(0.02)	(0.02)	(0.11)	(0.11)	(0.10)	(0.11)
Cuadrado	T	30	0.78	0.00	0.02	0.06	0.03	0.00	0.01	0.05	0.02	0.78	0.78	0.78	0.77
Cuadrado	T	30	(0.12)	(0.00)	(0.02)	(0.04)	(0.04)	(0.00)	(0.01)	(0.02)	(0.02)	(0.12)	(0.12)	(0.12)	(0.12)
Cuadrado	Tr	5	0.79	0.00	0.02	0.07	0.02	0.00	0.02	0.05	0.02	0.79	0.79	0.79	0.79
Cuadrado	Tr	5	(0.10)	(0.00)	(0.09)	(0.14)	(0.08)	(0.00)	(0.01)	(0.01)	(0.03)	(0.10)	(0.11)	(0.11)	(0.11)
Cuadrado	Tr	15	0.78	0.00	0.03	0.07	0.03	0.00	0.01	0.05	0.01	0.78	0.78	0.78	0.78
Cuadrado	Tr	15	(0.12)	(0.00)	(0.04)	(0.07)	(0.05)	(0.00)	(0.01)	(0.01)	(0.02)	(0.12)	(0.12)	(0.12)	(0.12)
Cuadrado	Tr	30	0.78	0.00	0.02	0.07	0.02	0.00	0.01	0.05	0.01	0.78	0.78	0.78	0.78
Cuadrado	Tr	30	(0.11)	(0.00)	(0.03)	(0.04)	(0.03)	(0.00)	(0.01)	(0.02)	(0.02)	(0.11)	(0.11)	(0.11)	(0.10)
Cuadrado	A	5	0.79	0.00	0.05	0.12	0.04	0.00	0.02	0.05	0.01	0.79	0.79	0.80	0.79
Cuadrado	A	5	(0.11)	(0.03)	(0.12)	(0.17)	(0.12)	(0.00)	(0.01)	(0.01)	(0.02)	(0.11)	(0.10)	(0.10)	(0.10)
Cuadrado	A	15	0.77	0.00	0.03	0.08	0.02	0.00	0.01	0.05	0.01	0.77	0.77	0.78	0.77
Cuadrado	A	15	(0.10)	(0.00)	(0.05)	(0.08)	(0.04)	(0.00)	(0.01)	(0.02)	(0.02)	(0.10)	(0.11)	(0.10)	(0.11)
Cuadrado	A	30	0.77	0.00	0.02	0.07	0.02	0.00	0.01	0.05	0.01	0.77	0.77	0.78	0.78
Cuadrado	A	30	(0.10)	(0.00)	(0.03)	(0.04)	(0.03)	(0.00)	(0.01)	(0.02)	(0.02)	(0.10)	(0.10)	(0.10)	(0.10)
Ajedrez	T	5	0.82	0.00	0.06	0.12	0.06	0.00	0.02	0.05	0.02	0.82	0.82	0.82	0.82
Ajedrez	T	5	(0.08)	(0.00)	(0.13)	(0.19)	(0.13)	(0.00)	(0.01)	(0.01)	(0.02)	(0.08)	(0.09)	(0.08)	(0.08)
Ajedrez	T	15	0.79	0.00	0.03	0.09	0.04	0.00	0.01	0.05	0.01	0.79	0.80	0.80	0.79
Ajedrez	T	15	(0.11)	(0.00)	(0.05)	(0.07)	(0.07)	(0.00)	(0.01)	(0.01)	(0.02)	(0.11)	(0.11)	(0.10)	(0.11)
Ajedrez	T	30	0.79	0.00	0.03	0.08	0.03	0.00	0.01	0.04	0.01	0.79	0.79	0.80	0.80
Ajedrez	T	30	(0.12)	(0.00)	(0.03)	(0.04)	(0.04)	(0.00)	(0.01)	(0.02)	(0.02)	(0.12)	(0.11)	(0.11)	(0.12)
Ajedrez	Tr	5	0.83	0.00	0.06	0.17	0.07	0.00	0.01	0.05	0.01	0.83	0.83	0.84	0.83
Ajedrez	Tr	5	(0.11)	(0.00)	(0.13)	(0.20)	(0.14)	(0.00)	(0.01)	(0.01)	(0.02)	(0.11)	(0.11)	(0.10)	(0.10)
Ajedrez	Tr	15	0.82	0.00	0.04	0.11	0.04	0.00	0.01	0.04	0.01	0.82	0.83	0.83	0.83
Ajedrez	Tr	15	(0.09)	(0.01)	(0.05)	(0.09)	(0.07)	(0.00)	(0.01)	(0.02)	(0.02)	(0.09)	(0.10)	(0.10)	(0.09)
Ajedrez	Tr	30	0.79	0.00	0.03	0.09	0.03	0.00	0.01	0.04	0.01	0.79	0.79	0.80	0.80
Ajedrez	Tr	30	(0.11)	(0.00)	(0.03)	(0.05)	(0.04)	(0.00)	(0.01)	(0.02)	(0.01)	(0.11)	(0.11)	(0.11)	(0.11)
Ajedrez	C	5	0.82	0.00	0.06	0.12	0.06	0.00	0.02	0.05	0.01	0.82	0.83	0.83	0.82
Ajedrez	C	5	(0.10)	(0.00)	(0.13)	(0.18)	(0.14)	(0.00)	(0.01)	(0.01)	(0.02)	(0.10)	(0.11)	(0.11)	(0.10)
Ajedrez	C	15	0.81	0.00	0.03	0.10	0.03	0.00	0.01	0.05	0.01	0.81	0.82	0.82	0.82
Ajedrez	C	15	(0.09)	(0.00)	(0.05)	(0.08)	(0.06)	(0.00)	(0.01)	(0.02)	(0.02)	(0.09)	(0.09)	(0.09)	(0.09)
Ajedrez	C	30	0.80	0.00	0.03	0.08	0.03	0.00	0.01	0.04	0.01	0.80	0.81	0.80	0.81
Ajedrez	C	30	(0.10)	(0.00)	(0.03)	(0.05)	(0.04)	(0.00)	(0.01)	(0.02)	(0.01)	(0.10)	(0.09)	(0.10)	(0.10)

**Tabla A-23.:** Promedio y desviación estándar (en paréntesis) de  $ABC$ , tasa de correcta detección ( $v_c$ ), tasa de falsa detección ( $f_c$ ) de cada método de identificación de influentes, en 100 replicaciones. Modelo de respuesta binaria,  $n = 70$ , bajo de contaminación de cambio forma - Tipo 1 en parámetro imagen.

$\beta_0$	$\gamma_c$ (%)	$RMSPE$	$v_c$				$f_c$				$RMSPE$ (sin influentes)			
			$DO$	$DO_b(0.995)$	$DO_b(0.95)$	$MS$	$DO$	$DO_b(0.995)$	$DO_b(0.95)$	$MS$	$DO$	$DO_b(0.995)$	$DO_b(0.95)$	$MS$
T	5	2.38	0.01	0.11	0.16	0.22	0.00	0.02	0.05	0.10	2.39	2.50	2.58	2.71
T	5	(0.45)	(0.07)	(0.21)	(0.26)	(0.28)	(0.01)	(0.01)	(0.02)	(0.06)	(0.45)	(0.47)	(0.51)	(0.53)
T	15	2.41	0.00	0.08	0.13	0.22	0.00	0.01	0.04	0.10	2.41	2.49	2.56	2.69
T	15	(0.52)	(0.02)	(0.08)	(0.09)	(0.15)	(0.00)	(0.01)	(0.02)	(0.07)	(0.52)	(0.53)	(0.58)	(0.60)
T	30	2.46	0.00	0.05	0.10	0.18	0.00	0.01	0.03	0.08	2.46	2.50	2.56	2.62
T	30	(0.43)	(0.00)	(0.04)	(0.05)	(0.10)	(0.00)	(0.01)	(0.02)	(0.06)	(0.43)	(0.44)	(0.45)	(0.47)
Triangulo	5	1.78	0.00	0.09	0.15	0.22	0.00	0.02	0.05	0.11	1.78	1.84	1.89	1.99
Triangulo	5	(0.38)	(0.00)	(0.19)	(0.25)	(0.30)	(0.00)	(0.01)	(0.02)	(0.07)	(0.38)	(0.36)	(0.37)	(0.38)
Triangulo	15	1.81	0.00	0.07	0.13	0.19	0.00	0.02	0.04	0.09	1.81	1.84	1.90	1.96
Triangulo	15	(0.33)	(0.02)	(0.08)	(0.11)	(0.17)	(0.00)	(0.01)	(0.02)	(0.06)	(0.33)	(0.35)	(0.36)	(0.35)
Triangulo	30	1.89	0.00	0.05	0.10	0.18	0.00	0.01	0.03	0.09	1.90	1.94	1.99	2.06
Triangulo	30	(0.32)	(0.01)	(0.03)	(0.05)	(0.11)	(0.00)	(0.02)	(0.02)	(0.06)	(0.32)	(0.33)	(0.34)	(0.38)
Cuadrado	5	1.93	0.01	0.11	0.17	0.26	0.00	0.02	0.05	0.11	1.93	2.05	2.14	2.29
Cuadrado	5	(0.41)	(0.07)	(0.21)	(0.26)	(0.32)	(0.00)	(0.01)	(0.02)	(0.06)	(0.41)	(0.43)	(0.47)	(0.50)
Cuadrado	15	2.07	0.01	0.08	0.15	0.24	0.00	0.01	0.04	0.10	2.08	2.13	2.18	2.34
Cuadrado	15	(0.43)	(0.03)	(0.08)	(0.10)	(0.16)	(0.00)	(0.01)	(0.02)	(0.06)	(0.42)	(0.43)	(0.44)	(0.45)
Cuadrado	30	2.01	0.00	0.05	0.11	0.19	0.00	0.01	0.03	0.09	2.01	2.08	2.14	2.28
Cuadrado	30	(0.42)	(0.02)	(0.04)	(0.05)	(0.11)	(0.00)	(0.02)	(0.02)	(0.07)	(0.42)	(0.42)	(0.43)	(0.46)
Ajedrez	5	2.52	0.00	0.01	0.07	0.10	0.00	0.02	0.05	0.10	2.52	2.65	2.77	2.88
Ajedrez	5	(0.49)	(0.00)	(0.09)	(0.17)	(0.22)	(0.00)	(0.01)	(0.01)	(0.06)	(0.50)	(0.54)	(0.59)	(0.65)
Ajedrez	15	2.69	0.00	0.01	0.05	0.09	0.00	0.02	0.05	0.10	2.70	2.83	2.95	3.16
Ajedrez	15	(0.50)	(0.00)	(0.05)	(0.08)	(0.12)	(0.00)	(0.01)	(0.02)	(0.06)	(0.50)	(0.52)	(0.52)	(0.53)
Ajedrez	30	2.53	0.00	0.02	0.05	0.09	0.00	0.02	0.05	0.10	2.53	2.65	2.77	2.89
Ajedrez	30	(0.55)	(0.01)	(0.03)	(0.05)	(0.08)	(0.01)	(0.02)	(0.03)	(0.07)	(0.54)	(0.56)	(0.61)	(0.63)

**Tabla A-24.:** Promedio y desviación estándar (en paréntesis) de  $RMSPE$ , tasa de correcta detección ( $v_c$ ), tasa de falsa detección ( $f_c$ ) de cada método de identificación de influentes, en 100 replicaciones. Modelo de respuesta continua,  $n = 50$ , bajo de contaminación de cambio forma - Tipo 2 en parámetro imagen (zoom 50 % de parámetro imagen original).



$\beta_0$	$\gamma_c(\%)$	$RMSPE$	$v_c$				$f_c$				$RMSPE$ (sin influentes)			
			$DO$	$DO_b(0.995)$	$DO_b(0.95)$	$MS$	$DO$	$DO_b(0.995)$	$DO_b(0.95)$	$MS$	$DO$	$DO_b(0.995)$	$DO_b(0.95)$	$MS$
T	5	2.17	0.00	0.09	0.21	0.30	0.00	0.01	0.05	0.12	2.18	2.22	2.31	2.49
T	5	(0.39)	(0.03)	(0.16)	(0.21)	(0.25)	(0.00)	(0.01)	(0.01)	(0.05)	(0.39)	(0.39)	(0.40)	(0.42)
T	15	2.31	0.00	0.05	0.14	0.23	0.00	0.01	0.04	0.10	2.31	2.35	2.43	2.54
T	15	(0.40)	(0.01)	(0.05)	(0.09)	(0.14)	(0.00)	(0.01)	(0.01)	(0.04)	(0.40)	(0.42)	(0.44)	(0.46)
T	30	2.33	0.00	0.04	0.09	0.18	0.00	0.01	0.04	0.10	2.33	2.35	2.42	2.49
T	30	(0.35)	(0.01)	(0.02)	(0.04)	(0.09)	(0.00)	(0.01)	(0.02)	(0.06)	(0.35)	(0.35)	(0.34)	(0.40)
Triangulo	5	1.71	0.00	0.08	0.15	0.27	0.00	0.01	0.05	0.13	1.71	1.75	1.81	1.91
Triangulo	5	(0.30)	(0.00)	(0.14)	(0.21)	(0.26)	(0.00)	(0.01)	(0.01)	(0.07)	(0.30)	(0.28)	(0.29)	(0.32)
Triangulo	15	1.73	0.00	0.06	0.13	0.22	0.00	0.01	0.04	0.11	1.73	1.76	1.81	1.90
Triangulo	15	(0.27)	(0.02)	(0.05)	(0.10)	(0.12)	(0.00)	(0.01)	(0.02)	(0.05)	(0.27)	(0.27)	(0.28)	(0.29)
Triangulo	30	1.78	0.00	0.04	0.09	0.19	0.00	0.01	0.04	0.10	1.78	1.80	1.86	1.93
Triangulo	30	(0.31)	(0.01)	(0.02)	(0.04)	(0.10)	(0.01)	(0.01)	(0.02)	(0.06)	(0.31)	(0.31)	(0.31)	(0.32)
Cuadrado	5	1.83	0.00	0.07	0.20	0.28	0.00	0.01	0.05	0.11	1.84	1.90	1.97	2.12
Cuadrado	5	(0.33)	(0.03)	(0.14)	(0.22)	(0.26)	(0.00)	(0.01)	(0.01)	(0.05)	(0.34)	(0.38)	(0.39)	(0.42)
Cuadrado	15	1.89	0.01	0.06	0.15	0.25	0.00	0.01	0.04	0.10	1.89	1.96	2.02	2.15
Cuadrado	15	(0.36)	(0.03)	(0.06)	(0.09)	(0.14)	(0.00)	(0.01)	(0.02)	(0.05)	(0.36)	(0.37)	(0.37)	(0.39)
Cuadrado	30	1.93	0.00	0.04	0.11	0.20	0.00	0.00	0.03	0.10	1.94	1.96	2.02	2.18
Cuadrado	30	(0.32)	(0.02)	(0.02)	(0.04)	(0.09)	(0.00)	(0.01)	(0.02)	(0.05)	(0.32)	(0.34)	(0.34)	(0.36)
Ajedrez	5	2.40	0.00	0.01	0.05	0.13	0.00	0.02	0.05	0.12	2.42	2.53	2.67	2.89
Ajedrez	5	(0.40)	(0.00)	(0.07)	(0.13)	(0.21)	(0.01)	(0.01)	(0.01)	(0.06)	(0.40)	(0.46)	(0.50)	(0.55)
Ajedrez	15	2.40	0.00	0.02	0.06	0.11	0.00	0.02	0.05	0.11	2.40	2.51	2.64	2.77
Ajedrez	15	(0.39)	(0.01)	(0.04)	(0.07)	(0.10)	(0.00)	(0.01)	(0.02)	(0.06)	(0.39)	(0.41)	(0.43)	(0.44)
Ajedrez	30	2.38	0.00	0.01	0.06	0.11	0.00	0.02	0.05	0.12	2.40	2.50	2.64	2.89
Ajedrez	30	(0.44)	(0.01)	(0.02)	(0.05)	(0.09)	(0.01)	(0.01)	(0.02)	(0.07)	(0.44)	(0.45)	(0.47)	(0.54)

**Tabla A-25.:** Promedio y desviación estándar (en paréntesis) de  $RMSPE$ , tasa de correcta detección ( $v_c$ ), tasa de falsa detección ( $f_c$ ) de cada método de identificación de influentes, en 100 replicaciones. Modelo de respuesta continua,  $n = 70$ , bajo de contaminación de cambio forma - Tipo 2 en parámetro imagen (zoom 50% de parámetro imagen original).

# Bibliografía

- L. Chien and T. Tsou. Robust influence diagnostics for generalized linear models with continuous responses. *Statistica Neerlandica*, 68:324–343, 2014.
- R. D. Cook. Detection of influential observation in linear regression. *Technometrics*, 19:15–18, 1977.
- W. Dai and M. G. Genton. Multivariate functional data visualization and outlier detection. *Journal of Computational and Graphical Statistics*, 27(4):923–934, 2018.
- W. Dai and M. G. Genton. Directional outlyingness for multivariate functional data. *Computational Statistics & Data Analysis*, 131:50–65, 2019.
- M. Febrero-Bande, P. Galeano, and W. González-Manteiga. Measures of influence for the functional linear model with scalar response. *Journal of Multivariate Analysis*, 101:327–339, 2010.
- C. Happ. Object-oriented software for functional data. *Journal of Statistical Software*, 93(5):1–38, 2020.
- R. R. Hocking. *Methods and applications of linear models: regression and the analysis of variance*. John Wiley & Sons, London, second edition, 2013.
- L. Horváth and P. Kokoszka. *Inference for Functional Data with Applications*. Springer, New York, first edition, 2012.
- A. E. Ivanescu. Function-on-function regression for two-dimensional functional data. *Communications in Statistics-Simulation and Computation*, 47:2656–2669, 2018.
- D.-H. Jang and C. M. Anderson-Cook. Firework Plots for Evaluating the Impact of Outliers and Influential Observations in Generalized Linear Models. *Quality Technology & Quantitative Management*, 12:423–436, 2015.
- P. Kokoszka and M. Reimherr. *Introduction to functional data analysis*. Chapman and Hall/CRC, 2017.
- R. Y. Liu. On a notion of data depth based on random simplices. *The Annals of Statistics*, pages 405–414, 1990.

- 
- P. McCullagh and J. A. Nelder. *Generalized Linear Models*. Chapman and Hall, London, second edition, 1989.
- R. Poldrack, J. Mumford, and T. Nichols. *Handbook of functional MRI data analysis*. Cambridge University Press, 2011.
- J. Ramsay and B. W. Silverman. *Functional Data Analysis*. Springer, New York, second edition, 2005.
- P. J. Rousseeuw and K. V. Driessen. A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, 41(3):212–223, 1999.
- R. J. Rousseeuw, P. and M. Hubert. A measure of directional outlyingness with applications to image data and video. *Journal of Computational and Graphical Statistics*, 27(2):345–359, 2018.
- X. Wang, H. Z. Zhu, and for the Alzheimer’s Disease Neuroimaging Initiative. Generalized Scalar-on-Image Regression Models via Total Variation. *Journal of the American Statistical Association*, 112:1156–1168, 2017.
- S. Yang and G. Berdine. The receiver operating characteristic (ROC) curve. *The Southwest Respiratory and Critical Care Chronicles*, 5(19):34–36, 2017.